Essays on exploding processes and covariance estimation

SCUOLA
NORMALE
SUPERIORE

Mboussa Anga Gael

Supervisor

Prof. Roberto Renò

University of Verona

This dissertation is submitted for the degree of

Doctor of Philosophy

April 2020

Summary

The Drift Burst Hypothesis postulates the existence of short-lived locally explosive trends in the price paths of financial assets (Christensen et al. 2016). Prominent examples are the recent US equity and treasury flash crashes. The first chapter of this work focuses on stochastic differential equations generating drift bursts. We elaborate on the model proposed by Cont and Wagalath (2013, 2016), where drift bursts arise endogenously due to distressed selling and feedback trading. In the original formulation of the model, interest lies on the impact of distressed selling for volatility, while we focus on the implications for the drift. The model re-enforces the point that drift bursts are a natural and expected outcome of the interaction amongst financial intermediaries. An interesting feature of the model is that both drift and volatility are inflated during a burst, as consistent with broad empirical evidence. In the proposed setup, the transaction price is affected by a random shock, mean reversion and a feedback effect. The model is initially cast in discrete-time, and as we move to the continuous-time limit-we show that local drift explosions can erupt under suitable conditions on the shape of the price impact function of feedback trading. We build drift bursts into the continuous-time Ito semi-martingale model in such a way that the fundamental arbitrage-free property is preserved. We finally study the existence and uniqueness of the solution of the proposed model. We provide suitable assumptions such that both drift and diffusion coefficients satisfy the conditions in Theorem 5.5.7 of Karatzas and Shreve (2012), which are necessary and sufficient conditions for the existence of the solution of exploding processes. Under these conditions, we prove that the SDE with exploding coefficients admits a weak solution which is unique in the sense of probability distribution and defined up to the exploding time. We also provide some example of the SDE with exploding coefficients (exploding models). We show those exploding models admit a weak solution which is unique in the sense of probability distribution and defined up to the exploding time. We discuss how to estimate the parameters of exploding models. Finally, we propose the dynamic Euler method that allows us to simulate more accurately the trajectories of those exploding models.

In the second chapter, we propose a non-parametric estimator of the exploding diffusion coefficient. The non-parametric estimator is constructed based on the theory of Nadaraya-Watson estimator. In our case, we include the non-parametric estimator of drift coefficient in the estimator of the spot volatility that helps us to improve the accuracy of the estimator when estimating the exploding diffusion coefficient. To test the unbiased nature of our proposed non-parametric estimator, we then compare it to the one introduced by Florens-Zmirou by mean of Monte Carlo simulations of the CEV model and the exploding models described above. Our results show that the correction is extremely beneficial around price explosions. Finally, we estimate the exploding diffusion coefficient from the E-mini S&P 500 flash crash prices of May 2010.

The third Chapter is devoted to covariance estimation for large samples. It's well-known that, in case of large datasets, the estimation of the integrated covariance matrix using robust estimators is computationally difficult or implies significant data reduction. For instance, the QMLE method of Corsi et al. (2015) and Shephard and Xiu (2017) is affected by curse of dimensionality, while the realized kernel estimator of Barndorff-Nielsen et al. (2011) discards a lot of data due to the use of refresh-time synchronization.

In our work, we tackle the problem of dimensionality reduction in the estimation of integrated covariances in presence of market microstructure effects (such as bid/ask bounce) and the so called non-synchronous trading. We propose to model intraday prices through a dynamic factor model where idiosyncratic errors incorporate microstructure effects. Estimation is performed by assuming a linear-Gaussian state-space representation and employing an Expectation-Maximization algorithm. The model provides a fully parametric setup where one can treat asynchronicity as a missing data problem, in a similar fashion to Corsi et al. (2015). The main advantages are that covariances can be estimated using the entire informational content of the data and the problem of model selection, namely how to choose the correct number of factors, is easily addressed through a likelihood-based criterion. To guarantee identification, we impose restrictions on the parameter space as suggested by Bańbura and Modugno (2014).

Extensive Monte Carlo simulations show the accuracy of the estimator in recovering the true covariance matrix in presence of stochastic volatility in the factors and asynchronous trading effects.

Empirically, we study the factor structure of intraday NYSE prices and perform an horse race based on out-of-sample minimum variance portfolios to compare our estimator to standard integrated covariance estimators. Results indicate that a factor structure emerges after the first few hours of the trading day and that the dynamic

factor specification does not lead to a significant increase of portfolio variances compared to the full model with no factor structure. We finally discuss the computational advantages of the proposed method.

Contents

List of Figures

List of Tables

Chapter 1

Modeling explosion of financial prices

1.1 Introduction

The drift burst hypothesis postulate the existence of short-lived locally explosive trends in the price path of the financial asset (Christensen et al. 2016). Prominent examples are the recent US equity and treasury flash crashes. In the rest of Chapter, the work focuses on stochastic differential equations generating drift bursts. We first build the model featuring drift burst. In the construction of the model, the drift burst arrives endogenously and they are a natural and expected outcome of the interaction amongst financial intermediaries. We show that the proposed drift burst model admits weak and strong solution. We provide the dynamic Euler algorithm to implement the drift burst model.

1.1.1 A model featuring the Drift Burst

To build the model featuring drift burst, we elaborate on the model proposed by Cont and Wagalath (2013) and Cont and Wagalath (2016), where drift bursts arise endogenously due to distressed selling and feedback trading. In the original formulation of the model, interest lies on the impact of distressed selling for volatility, while we focus on the implications for the drift. The model re-enforces the point that drift bursts are a natural and expected outcome of the interaction amongst financial intermediaries. An interesting feature of the model is that both drift and volatility are inflated during a burst, as consistent with broad empirical evidence in the Section 2 of (Christensen et al., 2016). In the proposed setup, the transaction price is affected by a random

shock, mean reversion and a feedback effect. The model is initially cast in discrete-time, and as we move to the continuous-time limit-we show that local drift explosions can erupt under suitable conditions on the shape of the price impact function of feedback trading. Let $t_i = i/n; i = 1,\dots,n$ be the discretization of the time interval $[0,T]$. Suppose that the trading takes place at discrete time points t_i and $\Delta = 1/n$ represents the time step between two consecutive transactions. Let $\tilde{X}_t$ denotes the efficient log-price. Then, we assume that the efficient log price starts at $\tilde{X}_0 = 0$ and evolves as a martingale:

$$\tilde{X}_{t_{i+1}} = \tilde{X}_{t_i} + \sigma_{\tilde{X}}\sqrt{\Delta}\tilde{\varepsilon}_{t_{i+1}}, \tag{1.1}$$

where $\tilde{\varepsilon}_{t_i} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$ represents the idiosyncratic component of efficient log-prices and $\sigma_{\tilde{X}} > 0$ controls the degree of underlying risk in the economy (or the volatility). Then, the transacted log-price X_t is assumed to follow:

$$\begin{aligned} X_{t_{i+1}} &= X_{t_i} + \sigma_X\sqrt{\Delta}\varepsilon_{t_{i+1}} + m(\bar{X}_{t_i} - X_{t_i})\Delta \\ &\quad + f(X_{t_i} + \sigma_X\sqrt{\Delta}\varepsilon_{t_{i+1}} + m(\bar{X}_{t_i} - X_{t_i})) - f(X_{t_i}) \end{aligned} \tag{1.2}$$

$\varepsilon_{t_i} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, m is an increasing function with $m(0) = 0$, and f is an increasing and concave function with $f(0) = 0$. The equation (1.2) can be interpreted in the following way:

- The first term $\sigma_X\sqrt{\Delta}\varepsilon_{t_{i+1}} \sim$ represents the contribution of uninformed traders, who do not posses any signal regarding the values of the efficient prices $\tilde{X}_t$. They trade randomly thereby moving the transaction price in a way that is independent of $\tilde{X}_t$.[1]

- The second term $m(\bar{X}_{t_i} - X_{t_i})\Delta \sim$ is the contribution of informed traders. The informed trader are the participants who do posses the knowledge (possibly via a noisy signal) of the efficient price $\tilde{X}_t$. They buy the security whatever the transaction price X_t is below the efficient price $\tilde{X}_t$ and sell the security whatever it above the efficient price $\tilde{X}_t$. Their price impact is modelled by an increasing function m with

[1] It is possible to allow for $\tilde{\varepsilon}_{t_i}$ to be correlated with ε_{t_i}. This does not change the line of thought behind the main result of this section, as the only difference is that in Eq. (1.3) B and W are correlated.

$m(0) = 0$. This implies that there is a tendency for X_t to revert towards $\tilde{X}_t$, which ensures market efficiency is preserved in the long run.

- The last term $f(\sigma_X \sqrt{\Delta}\varepsilon_{t_{i+1}} + m(\bar{X}_{t_i} - X_{t_i})) - f(X_{t_i}) \sim$ allows to capture the effect of feedback trading where directional price moves are re-enforced and magnified. This is achieved by letting the first derivative of the function f to be non-decreasing (i.e. $f' > 0$) with $f(0) = 0$.[2]

The presence of the feedback term in Eq. (1.2) can be justified by different mechanisms in the financial markets. For instance the liquidity provision theory of Grossman and Miller (1988). As highlighted by Eq. (1.2), with material one-sided liquidity demand (in our model this corresponds to a large value of the shock $\sigma_X \sqrt{\Delta}\varepsilon_{t_{i+1}}$) market makers that provide immediacy will increase their required compensation by moving X_t in the direction of demand. So for instance, with significant selling pressure, market makers will revise their quotes down and if the pressure persists they may do so more aggressively as their risk exposure builds up or the number of market makers diminishes. This can make the transaction price drop below the efficient price. At some stage, informed traders enter the market and drive the price back towards its fundamental level. This prediction is consistent with the CFTC and SEC (2010) report, where it was noticed that during the equity market flash crash, several participants reduced liquidity provision or withdrew entirely from the market in the midst of the turmoil. Subsequently, the market recovered back to its previous levels.

The literature on price formation is also abundant with models which imply feedback. Cont and Wagalath (2013, 2016) note that fire sales-i.e., the sudden deleveraging of large financial portfolios—can be self-reinforcing, leading to a downward spiral in asset prices. Gennotte and Leland (1990) develop a rational expectations model, inspired by the portfolio insurance strategies that were accused of exacerbating the 1987 stock market crash, where hedgers sell in a falling market to prevent further losses, thus creating a snowball effect that increases the initial price drop. Danielsson et al. (2012) note that feedback is present, when derivatives on the asset are traded, and option traders with short positions (i.e., negative "gamma") follow delta-hedging strategies to rehedge their risk. Barlevy and Veronesi (2003) show that market crashes can occur irrespective of strong fundamentals, if uninformed traders sell rationally at low prices in an attempt to extract information about the true value of the asset. In the model of forced liquidation and predatory trading by Brunnermeier and Pedersen (2005), strategic traders know that other market participants are in distress (e.g., due to margin calls, stop-loss orders, etc.). The

[2] While we assume that both m and f are time-homogeneous, in practice they can depend on different state variables of the market (e.g., volatility or liquidity).

informed trader is then inclined to sell the asset upfront, because he anticipates an opportunity to buy it back at better levels later on. Morris and Shin (2004) study a game-theoretic setup (inspired by bank run models), in which mutually reinforcing selling of short-term traders against a downward sloping demand curve from long-term traders can also result in market crashes.

The following result shows that the inclusion of a feedback term as in by equation (1.2) is capable of generating drift bursts in the continuous-time limit of X_t, provided that f is sufficiently concave:

Theorem 1 (Christensen et al., 2016)**.** *Assume that $E(\varepsilon^4) < \infty$ and f is three times differentiable, bounded and with bounded derivatives. Then, as $\Delta \to 0$, the dynamics in Eqs. (1.1)-(1.2) converge weakly to the processes:*

$$\begin{cases} d\bar{X}_t &= \sigma_{\bar{X}} dB_t, \\ dX_t &= \left([1+f'(X_t)]m(\bar{X}_t - X_t) + \frac{\sigma^{*2}}{2} f''(X_t)\right)dt + \sigma^*(1+f'(X_t))dW_t, \end{cases} \tag{1.3}$$

where B and W are independent standard Brownian motions.

The main aspect of the above Theorem is that an additional second-order effect-proportional to f'' appears in the drift term. This implies that the drift-to-volatility ratio of X_t is

$$\frac{\mu_t}{\sigma_t} = \frac{m(\bar{X}_t - X_t)}{\sigma^*} + \frac{\sigma^*}{2}\frac{f''(x)}{1+f'(x)}. \tag{1.4}$$

While the contribution of the informed trading is typically bounded, the second term, which is due to feedback trading, can diverge. As consistent with Section 2 in (Christensen et al., 2016), the proposed model can therefore exhibit explosions in both σ_t and μ_t in the domain of X_t via f', while $\frac{\mu_t}{\sigma_t} \to \infty$ is a possibility via f''. From the Eq. (1.4), we can see that the first term at right side of this equation is bounded and therefore it does not contribute to the explosion of the drift and difufusion coefficients. Thus, we can neglect the first term at right side of Eq. (1.4) and impose the drift-to-volatility ratio by solving the ODE in Eq. (1.4) by setting

$$\frac{\sigma^*}{2}\frac{f''(x)}{1+f'(x)} = g(x), \tag{1.5}$$

where the function $g(\cdot)$ models the drift-to-volatility ratio. We can solve the ODE (1.5) subject to a boundary condition $f(0) = f'(0) = 0$ and (1.5) yields

$$f'(x) = \exp\left(\frac{2}{\sigma^*}\int_0^x g(u)du\right) - 1 \quad \forall x \in \mathbb{R}. \tag{1.6}$$

The equation (1.6) can be integrated again to get f. The drift-to-volatility ratio $g(x)$ has implications for the feedback function. Note that $g(\cdot)$ is an explode function since the drift-to-volatility explodes. Substituting the equation (1.6) into (1.5) and SDE yields

$$\begin{cases} X_0 & = x_0 \\ dX_t & = \mu(X_t)dt + \sigma(X_t)dW_t, \\ \sigma(x) & = \sigma^* e^{\frac{2}{\sigma^*}\int_0^x g(u)du} \\ \mu(x) & = (m(x) + \sigma^* g(x))e^{\frac{2}{\sigma^*}\int_0^x g(u)du} \end{cases} \tag{1.7}$$

where $\mu : \mathbb{R} \longrightarrow \mathbb{R}$ and $\sigma : \mathbb{R} \longrightarrow \mathbb{R}^2$ are measurable. In the following subsection, we study the existence and uniqueness of solution of proposed stochastic differential equations.

1.2 One-dimensional time homogeneous SDE generating Drift bursts

In the above section we proposed a model, written in terms of a stochastic differential equation, generating a drift burst. In this section we focus on the studies of the properties of the proposed model. We investigate on the existence and uniqueness of strong (or weak) solution of the model generating drift burst. Through this section we assume that $(W_t)_{t\in[0,T]}$ is one-dimension Brownian motion on the filtered $(\Omega, \mathfrak{F}, (\mathfrak{F})_{t\geq 0}, P)$. We call a one-dimensional time homogeneous stochastic differential equation generating the drift burst an equation of

the form

$$\begin{cases} X_0 & = x_0 \\ dX_t & = \mu(X_t)dt + \sigma(X_t)dW_t. \\ \sigma(x) & = e^{\frac{2}{\sigma^*}\int_\cdot^x g(u)du} \\ \mu(x) & = (m(x) + \sigma^* g(x))e^{\frac{2}{\sigma^*}\int_\cdot^x g(u)du}, \end{cases} \tag{1.8}$$

where σ^* is a constant, $m(.) \in C^1$ and its first derivative are set be negative, e.i. $m'(.) < 0$ for all x, $X_0 - \mathfrak{F}_t$-adapted, g is an unknown function which is discontinuous in $\mathbb{R}$ and $\mu : \mathbb{R} \longrightarrow \mathbb{R}$ and $\sigma : \mathbb{R} \longrightarrow \mathbb{R}^2$ are measurable.

1.2.1 Literature review on SDEs

In literature, we know if the drift (μ) and diffusion (σ) coefficients satisfy the local Lipschitz and linear growth conditions, the SDE admits a solution which does not explode in the finite time (e.i.strong unique solution) for any given initial value. This solution exists and pathwise uniqueness (see, for example, Theorem 2.9, section 5.2 in Karatzas and Shreve 2012). Here are some good references for the SDE whose drift and diffusion coefficients satisfy the condition of linear growth, Engelbert and Schmidt (1991), Revuz and Yor (1999), Karatzas and Shreve (2012), Øksendal (2003). The linear growth condition may guarantee that the solution of the SDE does not explode at the finite time, but this condition may be too respective in practice. For instance, there exists the SDE whose coefficients dot not satisfy the linear growth condition, but still, satisfy the local Lipschitz condition. In this case, the solution of that equation may explode at the finite time (Liptser and Shiryayev 1989, Mao 2007).

About the pathwise uniqueness, we refer to the proof of Theorem IV.9.1 in Ikeda and Watanabe (1989) in which they used the (local) Lipschitz condition and Gronwall's inequality to demonstrated that L_2 distance between two solutions with the same initial value converges to zero. We note that the local Lipschitz condition can be relaxed or even weakened especially in the case of the one-dimensional SDEs. For instance, Yamada et al. (1971) consider Hölder condition instead of the (local) Lipschitz condition to prove the pathwise uniqueness solution of SDE. On the other hand, we should also note that there exist many situations, particularly in the case of one-dimensional SDE, in which the drift and diffusion coefficients and their corresponding derivatives not bounded. In this case, the existence and pathwise uniqueness of the solution of the SDE does not hold, but

perhaps the SDE may admit the local solution up to the exploding time. For example, Karatzas and Shreve (2012) show in Theorem 5.5.15 that, if $\sigma^2(\cdot) > 0$ and $\sigma^{-2}(\cdot)$, and $|\mu(\cdot)|\sigma^2(\cdot)$ are locally integrable for all point in $\mathbb{R}$, the SDE admits a weak solution which is unique in the sense of probability distribution and defined up until explosion time. The quoted literature deals with the problem of the existence and pathwise uniqueness of a strong (weak) solution of SDE whenever both drift and diffusion coefficients and their corresponding derivatives are bounded, or only the diffusion function is bounded. The main objective of the present Chapter is to study the properties of solutions of the one-dimension time-homogeneous stochastic differential equation generating drift burst. Note that the diffusion $\sigma(\cdot)$ and drift $\mu(\cdot)$ coefficients of the SDE in Eq. (1.8) depend on the unknown function $g(\cdot)$. As stated in the preview section 1.1.1, the drift-to-volatility g explodes at a single point of $\mathbb{R}$, and therefore the drift and diffusion coefficients do explode at a unique point of $\mathbb{R}$, with the drift coefficients explodes at a magnitude higher than that of the diffusion function. Comparing to the cases discussed above, we can see that the drift and diffusion coefficients of our SDE in Eq. (1.8) do not satisfy the regular conditions, local Lipschitz and linear growth conditions for every point of $\mathbb{R}$ since both coefficients explode at a unique point of $\mathbb{R}$. Thus, the SDE in Eq. (1.8) does not admit a strong solution since its coefficients do not satisfy the local Lipschitz and linear growth conditions. We aim to investigate the following questions:

(1) **Does the strong(weak) solution of SDE (1.8) exists? Is it unique?**

(2) **Do we have an example of strong(weak) solution of the SDE (1.8) for which both drift and volatility explode?**

In section 1.2.3, we answer to the question (1). Since the drift and diffusion coefficients of the SDE in Eq. (1.8) depend on the unknown function g. To investigate the existence and uniqueness of the strong (weak) solution of the SDE in Eq (1.8), we need to impose the standing assumption on the unknown function g. We assume that the function g explodes at single or multiple points of $\mathbb{R}$. Under this assumption, we show that the drift and diffusion coefficients of the SDE in Eq. (1.8) satisfy the same conditions of Theorem 5.5.15 of Karatzas and Shreve (2012). Thus, the SDE in Eq (1.8) admits a weak solution which is unique in the sense of the probability distribution and defined up until the explosion time (see Theorem 12 below). In section 1.2.4, we answer the question (2). We discuss some examples of the function g. We consider two different types of discontinuous functions of the $g(\cdot)$. Firstly, we assume that the function g explodes in two points of $\mathbb{R}$. Substituting this specific example of the function g in the expressions of the drift and diffusion coefficients of

the SDE (1.8), and we derive the parametric model with exploding coefficients. We then that the coefficients of the parametric model satisfy the condition in Theorem 12 below, and therefore, the solution of the model admits a weak solution which is unique in the sense of the probability distribution and defined up to explosion time. Second, we assume that the function g explodes at a single point of $\mathbb{R}$. Substituting the function g in the expressions of the drift and diffusion coefficients of the SDE (1.8), and we also derive another type of the parametric model with exploding coefficients. We also show the coefficients of the parametric model satisfy the condition of Theorem 12 and therefore, the parametric model admits the weak solution which is unique in the sense of the probability distribution and defined up to explosion time.

The rest of the section organizes as follows: First, we review some definitions of the SDE. Second, we answer the above questions on the existence and the uniqueness of the strong (weak) solution of SDE. Last, we provide some example of the SDE in which both drift and diffusion coefficients explode in the finite point.

1.2.2 Preliminaries on the Stochastic Differential Equation

In this subsection, we recall some known definitions such as the strong solution, weak solution, scale function, speed measure and explosion time. We then discuss the existence and uniqueness of the solution of the SDE (1.8) and some of its properties. The definitions, comments and theorems are taken from Veretennikov (1981), Revuz and Yor (1999), Mao and Yuan (2006), Karatzas and Shreve (2012).

Strong and Weak Solutions

We start with the definition of a strong solution of the d-dimension stochastic differential equation.

Definition 1. *(**Strong solution**) Suppose that $(\Omega, \mathfrak{F}, P)$ is a probability space with respected to the fixed $d-$dimensional Brownian motion W and independent initial value X_0 over this specific probability space. let $\{\mathfrak{F}_t\}_{t\geq 0}$ be a smallest filtration generated by the Wiener process W. A continuous process $X = \{X_t; t \geq 0\}$ with initial condition X_0 is called a strong solution of the stochastic differential equation if it satisfies the following properties:*

1) X is $\mathfrak{F}_t$-adapted

2) $P[X_0 = 0] = 1$,

3) $P\Big[\int_0^t\{|\mu(s,X_s)|+\sigma^2(s,X_s)\}<\infty\}\Big]$ *holds for all* $t\geq 0$,

4) the integral version of (1.8)

$$X_t = X_0 + \int_0^t \mu(s,X_s)ds + \int_0^t \sigma(s,X_s)dW_s, \tag{1.9}$$

holds almost surely.

The condition 1) in the above definition is the most crucial one, as indicates that the output of the solution X_t at time t is determined by equation and stochastic input of initial condition X_0, and the path of the Brownian motion $W=\{W_t;0\leq t<\infty\}$ up to that time. In the following section, we discuss the uniqueness of a strong solution of the SDE.

Definition 2. ***(Uniqueness of strong solution for SDE)*** *Suppose that both coefficients* (μ,σ) *are given. We say that* ***strong uniqueness*** *holds for the pair* (μ,σ)*, whatever* $\{W_t\}_{0\leq t<\infty}$ *is an* $d-$*dimension Brownian motion on some given probability space* $(\Omega,\mathfrak{F},P)$*,* X_0 *is an independent initial condition,* $\{\mathfrak{F}_t\}_{0\leq t<\infty}$ *is an augmented filtration, and if* X *and* $\bar{X}$ *are two strong solutions of the SDE with initial condition* X_0*, then* $P[X_t=\bar{X}_t;0\leq t\leq\infty]=1$ *holds.*

In the following examples we discuss some of the equations in which the strong uniqueness holds.

Example 1 (Karatzas and Shreve 2012)**.** *Consider the following one-dimension equation*

$$dX_t = \mu(t,X_t)dt + dW_t, \tag{1.10}$$

with a bounded, Borel-measurable function $\mu:\mathbb{R}^+\times\mathbb{R}\to\mathbb{R}$ *that is nonincreasing in the second variable. Then the strong uniqueness holds for the pair* $(\mu,1)$*. Suppose that* X^1 *and* X^2 *are two strong solutions of* (1.10) *on the same filtered probability space and with an initial value* X_0 *satisfying*

$$X_t^i = X_0 + \int_0^t \mu(s,X_s^i)ds + W_t; \quad 0\leq t<\infty \quad \textit{and} \quad i=1,2, \quad \textit{a.s.P.}$$

Then we can define the continuous process $\Delta_t^2 = X_t^1 - X_t^2$ *and observe that its differentiability with respect to t*

$$\frac{d}{dt}\Delta_t^2 = 2(X_t^1 - X_t^2)[\mu(t, X_t^1) - \mu(t, X_t^2)] \leq 0; \quad 0 \leq t < \infty, \quad a.s.\ P.$$

By integrating the above equation we get

$$\Delta_t^2 = 2\int_0^t (X_s^1 - X_s^2)[\mu(s, X_s^1) - \mu(s, X_s^2)]ds \leq 0; \quad 0 \leq t < \infty, \quad a.s.\ P.$$

Let now consider the case of deterministic differential equation. Suppose that the diffusion function σ is zero and then equation (1.22) reduces to

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds. \tag{1.11}$$

To verify if the solution of the above equation is unique, we must prove that the drift function satisfies the (local) Lipschitz condition in the space variable x, and is bounded on compact subset $\mathbb{R}^+ \times \mathbb{R}^d$ (see Theorem I.5.3 of Hale 1969). Now if the drift function fails to satisfy the (local) Lipschitz condition, and then the equation (1.11) may not be solvable. To clarify the argument, let us consider the following one-dimensional equation:

$$X_t = \int_0^t |X_s|^\delta ds. \tag{1.12}$$

If $\delta \in (0,1)$ then the equation (1.12) has family of solution of the form

$$X_t = \begin{cases} 0; & 0 \leq t \leq s \\ \left(\frac{t-s}{\alpha}\right)^\alpha; & s \leq t < \infty \end{cases}$$

with $\alpha = \frac{1}{1-\delta}$ and $0 \leq s \leq \infty$. If $\delta \geq 1$ then the equation (1.12) has one solution which is in fact $X_t \equiv 0$. From the above examples, we can see that the (local) Lipschitz condition is crucial when investigating the question of the existence and or uniqueness of the solution of the stochastic differential equation. The following Theorem of Karatzas and Shreve (2012) discusses the condition that guarantees the existence of a strong solution.

1.2 One-dimensional time homogeneous SDE generating Drift bursts

Theorem 2. *Suppose that the coefficients σ, μ satisfy the locally Lipschitz condition in the space variable x, e.i. for every integer $n \geq 0$ there exists a constant K_n such that for every $t \geq 0$, $\| x \| \leq n$ and $\| y \| \leq n$:*

$$\| \sigma(t,x) - \sigma(t,y) \| + \| \mu(t,x) - \mu(t,y) \| \leq K_n \| x - y \|, \tag{1.13}$$

holds. Then strong uniqueness holds for the SDE .

Proof. We refer to reader on the page: 288 in Karatzas and Shreve (2012), for the proof of the Theorem 2. □

We should point out that the (local) Lipschitz condition may not be sufficient to guarantee the global existence for the solution of the SDE. To verify the existence of the solution of SDE, we need to impose another strong condition. The following Theorem of Karatzas and Shreve 2012 discusses the condition for the existence of the solution of the SDE.

Theorem 3. *Suppose that the coefficients σ, μ satisfy the global Lipschitz and linear growth conditions on x, for every $0 \leq t < \infty, x \in \mathbb{R}^d, y \in \mathbb{R}^d$ and there exists a positive constants K such that:*

$$\| \sigma(t,x) - \sigma(t,y) \| + \| \mu(t,x) - \mu(t,y) \| \leq K \| x - y \|, \tag{1.14}$$

$$\| \sigma(t,x) \|^2 + \| \mu(t,x) \|^2 \leq K^2(1 + \| x \|^2). \tag{1.15}$$

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let x_0 be an $\mathbb{R}^d-$valued random point, independent of d-dimension Brownian motion $W = \{W_t, \mathfrak{F}_t^W; 0 \leq t < \infty\}$, and with finite second moment:

$$\mathbb{E} \| X_0 \|^2 < \infty. \tag{1.16}$$

Suppose that $\{\mathfrak{F}_t\}$ is an augmented filtration, and there exists a continuous, adapted process $X = \{X_t, \mathfrak{F}_t; 0 \leq t < \infty\}$ which is a strong solution of d-dimension SDE relative to W, with initial condition X_0. Moreover, this process is square-integrable: for every positive $T > 0$, there exists a constant C which dependent only on K and T, such that

$$\mathbb{E} \| X_t \|^2 \leq C(1 + \mathbb{E} \| x_0 \|^2) e^{Ct}; \quad 0 \leq t \leq T. \tag{1.17}$$

Proof. We refer the reader to the page : 291 of Karatzas and Shreve (2012) for the detail of a proof of Theorem 3. □

To prove the existence and the uniqueness of a strong solution of the SDE, we have to show the conditions of Theorems 2 and 3 hold. On the other hand, there are other types of SDE in which the conditions of Theorems 2 and 3 may not hold. In such a case, we can provide a weaker condition than those of the Theorems 2 and 3 to prove the existence and uniqueness of its solution. Below, we discuss the concept of the weak solution of the SDE.

Definition 3. *By a **weak solution** of the SDE (1.8) we mean a triplet* (X,W), $(\Omega,\mathfrak{F},P)$, $\{\mathfrak{F}_t\}$, *where*

i) $(\Omega,\mathfrak{F},P)$ *is a probability space equipped with the filtration* $\{\mathfrak{F}_t\}$ *that satisfies the usual conditions,*

ii) $W=\{W_t,\mathscr{F}_t;0\leqslant t<\infty\}$ *is a d-dimension* $(\mathfrak{F}_t)$*-Brownian motion on the probability space and* $X=\{X_t,\mathscr{F}_t,0\leqslant t<\infty\}$ *is continuous,* $(\mathfrak{F}_t)$*-adapted* $\mathbb{R}^d$*-value process;*

iii) the two last conditions of Definition (1) are fulfilled.

We refer to $P[X_0\in\Gamma],\Gamma\in\mathfrak{B}$ of X_0 where $\mathfrak{B}$ is a Borel set, as the **initial distribution** for the solution X.

Remark 1. *We note that any strong solution of the SDE is also a weak solution with an additional filtration* $\mathfrak{F}=\sigma(X_0)\vee\mathscr{F}_t,0\leq t<\infty$*, generated by the "driving Brownian motion "and" initial condition"* X_0*. In contrast, a weak solution does not imply a strong solution. We discuss below some examples of SDE whose solutions satisfy solutions satisfy the weak condition, but not the strong condition (Tanaka equation).*

Definition 4. *We say that the **pathwise uniqueness** holds for equation (1.8) if, for any two weak solutions* (X,W), $(\Omega,\mathfrak{F},P)$, $\{\mathfrak{F}_t\}$ *and* $(\bar{X},\bar{W})$, $(\bar{\Omega},\bar{\mathfrak{F}},\bar{P})$, $\{\bar{\mathfrak{F}}_t\}$ *on a common probability space with a common Brownian motion* W *with respect to both filtrations* $\{\mathfrak{F}_t\}$ *and* $\{\bar{\mathfrak{F}}_t\}$*, and with common initial value, i.e.*

$$P[X_0=\bar{X}_0]=1,$$

satisfy

$$P[X_t=\bar{X}_t;\forall 0\leq t\leq\infty]=1.$$

Definition 5. *We say that* ***uniqueness in the sense of probability law*** *holds for equation* (1.8) *whenever two weak solutions* (X,W), $(\Omega,\mathfrak{F},P)$, $\{\mathfrak{F}_t\}$ *and* $(\bar{X},\bar{W})$, $(\bar{\Omega},\bar{\mathfrak{F}},\bar{P})$, $\{\bar{\mathfrak{F}}_t\}$ *with the same initial distribution, i.e.*

$$P[X_0 \in \Gamma] = \bar{P}[\bar{X}_0 \in \Gamma]; \quad \forall \Gamma \in \mathfrak{B}(\mathbb{R}),$$

have the same law.

Remark 2. *Note that the existence of a weak solution does not guarantee the existence of a strong solution, and the uniqueness in the sense of probability law does not imply the pathwise uniqueness. However, the pathwise uniqueness implies the uniqueness in the sense of probability law (see Proposition* 3.20 *in Karatzas and Shreve 2012).*

Let us discuss the following Tanaka equation in order to clarify the remarks 1 and 2.

Example 2. *Let* X_t *be a one-dimensional equation given by*

$$X_t = \int_0^t sgn(X_s)dW_s; \quad 0 \leq t < \infty, \tag{1.18}$$

where

$$sgn(x) = \begin{cases} 1; & x > 0 \\ -1; & x \leq 0. \end{cases}$$

We observe that the conditions of Theorems (3) *and* (2) *are not fulfilled, therefore the equation* (1.18) *has no strong solution. We also observe that the pathwise uniqueness does not hold for equation* (1.18) *since both* X *and* $-X$ *are at the same time its solutions. Suppose that a triplet* (X,W), $(\Omega,\mathfrak{F},P)$, $\{\mathfrak{F}_t\}$ *is a weak solution of the equation*(1.18)*, then the process* $X = \{X_t,\mathfrak{F}_t; 0 \leq t < \infty\}$ *is a continuous, square-integral martingale with* $\langle X\rangle_t = \int_0^t sgn^2(X_s)ds = t$*. Therefore,* X *is a Brownian motion and thus uniqueness in the sense of probability law holds. For more detail about the discussion of this example we refer the reader on the page* 302 *of Karatzas and Shreve (2012).*

In the above discussion, we focused on the case of the multidimensional stochastic differential equation. However, the model under consideration is a one-dimensional case. It is, therefore, reasonable to discuss the conditions of the existence and uniqueness of the strong (weak) solution for the one-dimension SDE case.

One-dimension time homogeneous SDE case with explosion

In this subsequent section, we study the properties of the one-dimension time-homogeneous stochastic differential equation (1.8). We discuss the conditions of the existence and uniqueness of the strong (weak) solution of a one-dimension time-homogeneous stochastic differential equation. From the Theorems 2 and 3 we knew to prove the existence and uniqueness of the (global) strong solution of d-dimension SDE, we have to show that the pair (μ, σ) satisfy the regularities conditions such as the (local) Lipschitz condition and linear growth condition. However, in the case of a one-dimension, we only need to show one of the coefficients (drift coefficient or diffusion coefficient) satisfy the regular conditions.

Theorem 4. *(Yamada et al. 1971)*

Suppose that the pair (σ, μ) *of the SDE* (1.8) *satisfy the following conditions*

$$|\sigma(t,x) - \sigma(t,y)| \leq h(x-y) \tag{1.19}$$

$$|\mu(t,x) - \mu(t,y)| \leq K|x-y|. \tag{1.20}$$

For every $0 \leq t < \infty$ *and* $x, y \in \mathbb{R}$. *Here* K *is a positive constant and a function* $h : \mathbb{R}^+ \to \mathbb{R}^+$ *is a strictly increasing with* $h(0) = 0$ *and*

$$\int_0^{\varepsilon} h^{-2}(s)ds = \infty, \quad \forall \varepsilon > 0. \tag{1.21}$$

Then strong uniqueness holds for equation (1.8).

From the Theorem 4, we can see the pair (μ, σ) do not satisfy the linear growth condition. We also see that the solution of the SDE appears to explode in the finite time. Thus, if the coefficients of the SDE satisfy the conditions of Theorem (1.8) and its solution may explode in the finite time. Now if the coefficients of the SDE fail to satisfy the conditions in Theorem 4, in this case, the strong uniqueness does not hold as well.

Let us now discuss the conditions that guarantee the weak solution of one-dimensional of SDE (1.8). Thus, the following definition of the weak solution up to an explosion time (see also Karatzas and Shreve 2012).

Definition 6. *Let $(\Omega,\mathfrak{F},P)$ be a probability space and let $\{\mathfrak{F}_t\}$ be a filtration of sub-σ-fields of $\mathfrak{F}$ satisfying the usual conditions. Let $W=\{W_t,\mathscr{F}_t;t\in[0,\infty)\}$ be a standard one dimension $\mathfrak{F}_t$-Brownian motion. By a **weak solution up to an explosion time** of the SDE (1.8) we mean an $\mathbb{R}\cup\{\pm\infty\}$-valued and continuous $\mathscr{F}$-adapted process $X=\{X_t,\mathscr{F}_t,t\geq 0\}$ such that the equation*

$$X_{t\wedge S_n}=X_0+\int_0^{t\wedge S_n}\mu(X_s)ds+\int_0^{t\wedge S_n}\sigma(X_s)dW_s, \tag{1.22}$$

holds for all $n\geq 0$ and $0\leqslant t<\infty$ a.s., where the absolute value of the initial value is finite almost surely, e.i. $|X_0|<\infty$ a.s. and $S_n:=\inf\{t\geq 0;|X_t|\geq n\}$ is the first exit time from the closed ball $B(n):=\{x\in\mathbb{R}:|x|\leq n\}$. The following limit

$$S=\lim_{n\to\infty}S_n \tag{1.23}$$

is called the exploding time for process X.

A process X is said to be explosive if $P[S<\infty]=1$ and it said to be non-explosive or continuous if $P[S=\infty]=1$. Now, if $P[S=\infty]=1$ then Definition 6 reduces to Definition (1) for a weak solution to equation (1.8). As we are seeking for the conditions that guarantee the existence and strong (weak) uniqueness of solution X of the one-dimension case, then we begin with a discussion of the one-dimension SDE without drift.

Equation without drift

Suppose that $W=\{W_t,;t\in[0,\infty)\}$ is a one-dimension Brownian motion with corresponding augmented filtration $\{\mathscr{F}_t\}_{t\geq 0}$ on some probability space $(\Omega,\mathfrak{F},P)$ and with an independent random variable ξ with initial distribution ς. Suppose that $I(\sigma)$ is a closed set denoted by

$$I(\sigma)=\left\{x\in\mathbb{R};\int_{-\varepsilon}^{\varepsilon}\sigma^{-2}(x+y)dy=\infty,\forall\varepsilon>0\right\}, \tag{1.24}$$

and then we define

$$I(\sigma) \stackrel{\Delta}{=} \inf\{s \geq 0; \xi + W_s \in I(\sigma)\}. \tag{1.25}$$

Let us now consider the equation (1.8) with a drift function is identically zero, e.i. $\mu \equiv 0$, that is

$$dX_t = \sigma(X_t)dW_t. \tag{1.26}$$

To prove the existence and strong uniqueness of a non-exploding solution of (1.26) holds, we rely on the following results of Engelbert and Schmidt (1985).

Theorem 5. ***(Engelbert and Schmidt 1985)*** *The one-dimension SDE (1.8) with zero drift has a non-exploding weak solution for every initial distribution ς if only if*

$$I(\sigma) \subseteq Z(\sigma), \tag{1.27}$$

where

$$Z(\sigma) = \{x \in \mathbb{R}; \sigma(x) = 0\}.$$

The condition (1.27) is equivalent to, if

$$\int_{-\varepsilon}^{\varepsilon} \sigma^{-2}(x+y)dy = \infty, \forall \varepsilon > 0 \Rightarrow \sigma(x) = 0. \tag{1.28}$$

Proof. We refer the reader to the page 332 of Karatzas and Shreve (2012) for more details a bout the proof of Theorem (5). □

Note that every continuous diffusion function σ satisfies the condition (1.27). However, only some special types of the discontinuous diffusion coefficient satisfy the condition (1.27). For example, if $\sigma(x) = sign(x)$ we can prove that the condition (1.27) holds. However, if $\sigma(x) = 1_0(x)$, we can see that a a condition (1.27) does not hold. The condition (1.27) in Theorem (5) guarantees the existence of solution of the SDE (1.26). On the other hand, we also need a sufficient condition to guarantees the uniqueness of the solution of the SDE (1.26).

Thus, in the following Theorem, we discuss the condition that guarantees the uniqueness of the solution of the SDE.

Theorem 6. ***(Engelbert and Schmidt 1985)*** *The one-dimension SDE (1.8) with zero drift has a unique solution in sense of the probability in law for every initial distribution ς if only if*

$$I(\sigma) = Z(\sigma). \tag{1.29}$$

Proof. The details of proof of Theorem 5 can be found in Karatzas and Shreve (2012), page 335. □

We note that in order to prove a condition (1.29), one should prove the both inclusions $I(\sigma) \subseteq Z(\sigma)$ and $I(\sigma) \supseteq Z(\sigma)$. The first inclusion $I(\sigma) \subseteq Z(\sigma)$ is basically a condition (1.27) of Theorem (5) which is a sufficient condition for the existence of (1.26). We can say that the reverse inclusion $I(\sigma) \supseteq Z(\sigma)$ is then a sufficient for uniqueness. The condition of Theorem (5) and (6) guarantee the existence and uniqueness of weak solution for the non-explosive solution of (1.26). We also need to verify if the existence and uniqueness of the strong, non-explosion solution of (1.26) holds. (Karatzas and Shreve, 2012) shown that pathwise uniqueness and the existence of the weak solution imply the existence and uniqueness of the strong solution. Therefore, to verify if the existence and uniqueness of the strong for the non-explosive solution of (1.26) holds, we only need to prove that the solution of (1.26) fulfils the conditions of pathwise uniqueness properties. Thus, the following Theorem:

Theorem 7. ***(Karatzas and Shreve 2012)***

Suppose that there exist functions $f : \mathbb{R} \to [0,\infty)$ and $h : [0,\infty) \to [0,\infty)$ such that

(1) the quotient $(f * \sigma^{-1})^2$ is locally integral at every $x \in I(\sigma)^c$; e.i. there exists $\varepsilon > 0$ such that

$$\int_{x-\varepsilon}^{x+\varepsilon} (f(y)\sigma^{-1}(y))^2 dy < \infty; \tag{1.30}$$

(2) h is strictly increasing with $h(0) = 0$ and fulfils (1.21);

(3) there exists a constant $a > 0$ such that

$$|\sigma(x+y) - \sigma(x)| \leq f(x)h(|y|); \quad \forall x \in \mathbb{R}, y \in [-a,a]. \tag{1.31}$$

Under these conditions, the pathwise uniqueness of the SDE (1.26) *holds.*

Remark 3. *From Theorem 7, we observe that if the function f is locally bounded, then a condition (1) of Theorem 7 follows directly from the definition $I(\sigma)$. Suppose that $h(y) = y^{\alpha}, \alpha \geq \frac{1}{2}$ and then a condition (3) implies that a diffusion coefficient σ is locally hölder-continuous with exponent at least* $(1/2)$.

Equation with drift

Let turn back our attention to the SDE in Eq. (1.8) with non-zero drift coefficient. To treat the general case of the SDE with non-zero drift, we only need to transform the SDE (1.8) with a drift coefficient to the case without drift which is already studied.To do so, we need to remove the drift coefficient of the SDE (1.8). To remove the drift coefficient of the SDE (1.8), we rely on the method of removal of the drift of Karatzas and Shreve (2012) (see also Zvonkin 1974 and Leobacher et al. 2015). To apply the method of removal of drift, we require the following assumptions of **nondegeneracy** and **local integrability**:

(A) $\quad \sigma^2(x) > 0; \quad \forall x \in \mathbb{R},$

(A1) $\quad \forall x \in \mathbb{R}, \exists x > 0$ such that $\int_{x-\varepsilon}^{x+\varepsilon} \frac{|\mu(y)|}{\sigma^2(y)} dy < \infty,$

to hold. Under the assumptions (A) and (A1) we known that the existence of solution X of the SDE (1.8) hold. Let $I = (l, r); -\infty \leq l < r \leq \infty$ be a subset of $\mathbb{R}$ and let $H : \mathscr{I} \longrightarrow \mathbb{R}$ be a smooth function; and we then define a new function

$$Y_t = H(X_t). \tag{1.32}$$

Since a function H is smooth and by applying Itô's formula on (1.32) with respect to (1.8), we get

$$dY_t = \underbrace{\left(\mu(X_t) \frac{\partial H(X_t)}{\partial X_t} + \frac{1}{2} \sigma(X_t)^2 \frac{\partial^2 H(X_t)}{\partial X_t^2} \right)}_{\mathscr{A}(H(X_t))} dt + \sigma(X_t) \frac{\partial H(X_t)}{\partial X_t} dW(t). \tag{1.33}$$

By imposing the operator $\mathscr{A}(H(x)) = \mu(x)\frac{\partial H(x)}{\partial x} + \frac{1}{2}\sigma(x)^2\frac{\partial^2 h(x)}{\partial x^2} = 0$ and thus a function H is obtained by solving the following equality:

$$\frac{d}{dx}\ln\left|\frac{\partial H(x)}{\partial x}\right| = -\frac{2|\mu(x)|}{\sigma(x)^2}.$$

As we are looking for a convenient function H, one can assume that $H' > 0$. Thanks to the assumption (A1) and a simple calculations leads to

$$H(x) = \int_c^x \exp\left(-\int_a^y \frac{2\mu(z)}{\sigma(z)^2}dz\right)dy, \tag{1.34}$$

where c, a are two positive constants. The function H is differentiated on $\mathscr{I}$ and its derivative is given by

$$H'(x) = \exp\left(-\int_a^x \frac{2\mu(x)}{\sigma(x)^2}dx\right), \quad \forall x, a \in \mathscr{I}, \tag{1.35}$$

increases for all $x \in \mathscr{I} = (l, r)$. From the definition of **scale function** see Revuz and Yor (1999), we observe that $H(\cdot)$ is exactly the *scale function* of diffusion $X(\cdot)$ with $c = a$. Since $H' > 0$, by inverse function theorem (see, Theorem 9.24 in Rudin (1976)), and we say that the function H is globally invertible on $\mathscr{I}$ with the inverse function given by $\mathfrak{q} = H^{-1}$. The inverse function maps from $H(\mathscr{I})$ onto the interval $\mathscr{J} = (\lambda, \nu)$ with endpoint $\lambda = H(l+)$ and $\nu = H(r-)$. Since $\mathscr{A}(H(x)) = 0$, we denote by Y , the solution of the SDE (1.8) without drift:

$$dY_t = \hat{\sigma}(Y_t)dW(t), \tag{1.36}$$

$$\hat{\sigma}(Y_t) = \sigma(\mathfrak{q}(Y_t))H'(\mathfrak{q}(Y_t)), \tag{1.37}$$

where W is one-dimensional Brownian motion. The conditions of the existence and uniqueness of the strong (weak) non-explosive solution of (1.36) are established in the following Theorem of Karatzas and Shreve (2012):

Theorem 8. ***(Removal of Drift)***

Assume that the conditions (A) and (A1) hold and let H be a scale function given in the equation (1.34) with $c = a = \varepsilon$. A process $X = \{X_t, \mathfrak{F}_t; 0 \le t < \infty\}$ is a strong (or weak) solution of Eq. (1.8) if only if the process

$Y = \{Y_t = H(X_t), \mathfrak{F}_t; 0 \leq t < \infty\}$ *is a strong (or weak) solution of*

$$Y_t = Y_0 + \int_0^t \hat{\sigma}(Y_s) dW(s); \quad 0 \leq t < \infty, \tag{1.38}$$

where

$$H(-\infty) < Y_0 < H(\infty) \quad a.s, \tag{1.39}$$

$$\hat{\sigma}(y) = \begin{cases} \sigma(\mathfrak{q}(y))H'(\mathfrak{q}(y)), & H(-\infty) < y < H(\infty), \\ 0 & otherwise \end{cases} \tag{1.40}$$

The solution X of the SDE (1.8) *may explode in finite time, but the solution Y of the* (1.36) *does not.*

Proof. We refer the reader on the page: 340 of Karatzas and Shreve (2012) for details. □

From the Theorem (8), we can see that the solution X may explode in finite time. To guarantee the non-explosion of the solution X of the SDE, we require that the condition $H(\pm) = \pm\infty$ must be satisfied. If this condition holds and then the solution X will not explode in finite time. From the Theorem (8), we now have the conditions that guarantee the existence and uniqueness of strong (weak) of the (no)-exploding solution of the SDE with non-zero drift. Below, we discuss the conditions that guarantee the existence and uniqueness of the weak solution up to explosion time in the sense of probability law of the SDE with the non-zero drift.

Theorem 9. ***(Karatzas and Shreve 2012)***

Suppose that σ^{-2} *is locally integral at every point in* $\mathbb{R}$ *and the conditions (A) and (A1) are fulfilled. Then the solution X of the SDE* (1.8) *with initial distribution* ξ *admits a weak solution up to the explosion time which is unique in the sense of the probability law.*

Proof. We refer the reader on the page: 341 of Karatzas and Shreve (2012) for details. □

The Theorem 9 is important in our case because the result of the next section is established based on the conditions of Theorem 9. Note that the endpoint of $\mathscr{I} = (r, l)$ are absorbing for the process $X(\cdot)$, which means

once the process X hits the endpoint of $\mathscr{I}$, it will stay there forever. To verify whatever the explosion occurs, we rely on the well-known *Feller's test of explosion*, (see Theorem below (Karatzas and Shreve 2012):

Theorem 10 (Feller test for Explosion)**.**

Let assume that the coefficients $\sigma : I \to \mathbb{R}, \mu : I \to \mathbb{R}$ *satisfy the following conditions:*

(A)' $\quad \sigma^2(x) > 0; \quad \forall x \in I,$

(A1)' $\quad \forall x \in I, \exists x > 0$ *such that* $\int_{x-\varepsilon}^{x+\varepsilon} \frac{1+|\mu(y)|}{\sigma^2(y)} dy < \infty,$

and let (X,W), $(\Omega,\mathfrak{F},P)$, $\{\mathfrak{F}_t\}$ *be a weak solution of the SDE* (1.8) *in* I *with some nonrandom initial condition* $x_0 \in I$. *Then* $P[S=\infty]=1$ *or* $P[S=\infty]<1$ *holds or not if only if*

$$\omega_e(r-) = \omega_e(l+) = \infty, \tag{1.41}$$

where

$$\omega_e(x) = \int_e^x H'(z) \left(\int_e^z \frac{2dy}{H'(y)\sigma^2(y)} \right) dz, \quad e \in I. \tag{1.42}$$

Remark 4. *Note that a constant e does not have any influence on the finiteness or non-finiteness of* $\omega_e(r-)$ *and* $\omega_e(l+)$.

1.2.3 Existence and Uniqueness of solution of the SDE generating Drift burst

In this subsection we now attempt to answer question (1) asked in a Section 1.1.1. The results proposed here are constructed based on the results discussed in the subsections (1.2.2) and (1.2.2). The model is

$$\begin{cases} X_0 & = x \\ dX_t & = \mu(X_t)dt + \sigma(X_t)dW_t. \end{cases} \tag{1.43}$$

with

$$\mu(x) = (m(x) + \sigma^* g(x)) e^{\frac{2}{\sigma^*}\int_c^x g(u)du} \tag{1.44}$$

$$\sigma(x) = e^{\frac{2}{\sigma^*}\int_{\bar{X}}^x g(u)du}, \tag{1.45}$$

where $m(.) \in C^1$, $m'(.) < 0$ for all $x \in \mathbb{R}$ and g is an unknown function (or a drift-to-volatility ratio). $(W_t)_{t\in[0,T]}$ is one-dimension Brownian motion in the filtered probability space $(\Omega, \mathscr{P}, (\mathscr{F}_t)_{t\geq 0}, \mathscr{F})$, x is square integrable, σ^* is strictly positive constant and $\bar{X}$ is a point before the flash crash starts. We know that $m(\cdot)$ is bounded and therefore the explosion of drift function does not depend on $m(\cdot)$. Thus by assuming $m(x) = 0$ for all x and the results in the rest of this work will not be affected. In the rest of this work, we now set $m(x) = 0$ for all x , and threfore the drift coefficient reduces to the following form:

$$\mu(x) = \sigma^*(x) e^{\frac{2}{\sigma^*}\int_c^x g(u)du}. \tag{1.46}$$

Thus, the diffusion and drift functions of Eq. (1.43) depend on an unknown function g. To provide the conditions that guarantee the existence and uniqueness of strong (weak) (non-)exploding solution of SDE (1.43), we first need to discuss the properties of g. Note that the function g is a discontinuous function and there exist many types of discontinuous functions such as the jump discontinuous, the infinite discontinuous, the removal discontinuous, and so forth. However, we knew that a function g converges to infinity at exploding point, and therefore g must be of the class of infinity discontinuity.

Assumption 1. *Suppose that the function g explodes at single point in $\mathbb{R}$. Suppose that x_e is an exploding point (or singular point) in $\mathbb{R}$ and then $g(x)$ converges to infinity as x approaches x_e.*

Theorem 11.

Suppose that assumption 1 holds. Let $I = (l, r); -\infty \leq l < r \leq \infty$ be a subset of $\mathbb{R}$ and let assume that the coefficients $\sigma : I \to \mathbb{R}$, and $\mu : I \to \mathbb{R}$ are measurable and satisfy the following conditions:

(1) $\sigma^2(x) > 0; \quad \forall x \in \mathbb{R},$

(2)

$$\int_{\mathfrak{K}} \left(\frac{1}{\sigma^2(y)} + \frac{|\mu(y)|}{\sigma^2(y)} \right) dy < \infty \quad \textit{for every compact set } \mathfrak{K} \subset I. \tag{1.47}$$

Proof. We start by verifying a condition (1). The condition, (1) implies that the diffusion coefficient is non-degenerate, and that means the diffusion coefficient is bounded away from zero and infinity. Let us now prove that $\sigma^2(x) > 0, \forall x \in \mathbb{R}$ and note that $\sigma(x) = \sigma^* e^{\frac{2}{\sigma^*}\int_{\bar{X}}^{x} g(y)dy}$. We prove a condition (1) by contraction. Assume that a condition (1) is false, that means for all $x \in \mathbb{R}$:

$$\sigma^2(x) = \sigma^{*2} e^{\frac{4}{\sigma^*}\int_{\bar{X}}^{x} g(y)dy} < 0. \tag{1.48}$$

Equality (1.48) implies that

$$\begin{aligned} &\frac{4}{\sigma^*}\int_{\bar{X}}^{x} g(y)dy < -\infty \\ &\Rightarrow -\frac{4}{\sigma^*}\int_{\bar{X}}^{x} g(y)dy > +\infty. \end{aligned} \tag{1.49}$$

The last inequality (1.52) cannot be true since this integral cannot be strictly greater than plus infinity even at exploding any $x = x_e$. Hence a condition (1) is true. We first must show that $|\mu(y)|\sigma^{-2}$ is locally integral at every point $x_0 \in (l, r)$. At such a point, choose $\varepsilon > 0$ so that

$$l < x_0 - \varepsilon \leq x_e < x_0 + \varepsilon < r,$$

and write

$$\int_{x_0-\varepsilon}^{x_0+\varepsilon} \left(\frac{|\mu(y)|}{\sigma^2(y)} \right) dy = \frac{1}{\sigma^*} \int_{x_0-\varepsilon}^{x_0+\varepsilon} |g(z)| e^{-\frac{2}{\sigma^*}\int_{\bar{X}}^{z} g(y)dy} dz. \tag{1.50}$$

Suppose that $g(x) > 0$ for every $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$ and then the integral (1.50) yields

$$\frac{1}{\sigma^*}\int_{x_0-\varepsilon}^{x_0+\varepsilon} g(z)e^{-\frac{2}{\sigma^*}\int_{\bar{X}}^{z} g(y)dy}dz = -\frac{1}{2\sigma^*}e^{-\frac{2}{\sigma^*}\int_{\bar{X}}^{x_0+\varepsilon} g(y)dy} + \frac{1}{2\sigma^*}e^{-\frac{2}{\sigma^*}\int_{\bar{X}}^{x_0-\varepsilon} g(y)dy}. \tag{1.51}$$

According to a condition (1), we have

$$\begin{aligned} &\sigma^{2*}e^{\frac{4}{\sigma^*}\int_{\bar{X}}^{x} g(y)dy} > 0 \\ &\Rightarrow \frac{4}{\sigma^*}\int_{\bar{X}}^{x} g(y)dy > -\infty \\ &\Rightarrow -\frac{2}{\sigma^*}\int_{\bar{X}}^{x} g(y)dy < +\infty \end{aligned} \tag{1.52}$$

Using Eq. (1.52) and for every $\varepsilon > 0$, and we obtain

$$\begin{aligned} &\Rightarrow -2\int_{\bar{X}}^{x_0+\varepsilon} g(y)dy < +\infty \\ &\Rightarrow e^{-2\int_{\bar{X}}^{x_0+\varepsilon} g(y)dy} < +\infty. \end{aligned} \tag{1.53}$$

Similarly we also obtain

$$\Rightarrow e^{-2\int_{\bar{X}}^{x_0-\varepsilon} g(y)dy} < +\infty. \tag{1.54}$$

From (1.53) and (1.54) we can say that the right side of the integral (1.51) is finite even around the exploding x_e. Assume now $g(x) < 0$ for every $x \in (x_0-\varepsilon, x_0+\varepsilon)$ and then the integral (1.50) yields

$$-\frac{1}{\sigma^*}\int_{x_0-\varepsilon}^{x_0+\varepsilon} g(z)e^{-\frac{2}{\sigma^*}\int_{\bar{X}}^{z} g(y)dy}dz = \frac{1}{2\sigma^*}e^{-\frac{2}{\sigma^*}\int_{\bar{X}}^{x_0+\varepsilon} g(y)dy} - \frac{1}{2\sigma^*}e^{-\frac{2}{\sigma^*}\int_{\bar{X}}^{x_0-\varepsilon} g(y)dy}. \tag{1.55}$$

In the same way as in the previous proof we also see that the right side of (1.55) is finite and hence we conclude that $|\mu(y)|\sigma^{-2}$ is locally integrable.

We now need to show that σ^{-2} is locally integrable at every point $x_0 \in (l,r)$. At such point and we also choose $\varepsilon > 0$ so that

$$l < x_0 - \varepsilon \leq x_e < x_0 + \varepsilon < r,$$

and then write

$$\int_{x_0-\varepsilon}^{x_0+\varepsilon} \sigma^2(y)dy = \sigma^{2*} \int_{x_0-\varepsilon}^{x_0+\varepsilon} e^{-\frac{4}{\sigma^*}\int_{\bar{X}}^{z} g(y)dy} dz. \tag{1.56}$$

Now using the fact that $\sigma^2(x) > 0$ for every $x \in (x_0-\varepsilon, x_0+\varepsilon)$ and apply the same idea as in the previous proof, and then we obtain

$$\begin{aligned} & e^{-\frac{4}{\sigma^*}\int_{\bar{X}}^{x_0+\varepsilon} g(y)dy} < +\infty. \\ \Rightarrow & \int_{x_0-\varepsilon}^{x_0+\varepsilon} e^{-\frac{4}{\sigma^*}\int_{\bar{X}}^{z} g(y)dy} dz < \infty. \end{aligned} \tag{1.57}$$

Thus, we conclude that σ^{-2} is locally integrable. □

Thus, we conclude that σ^{-2} is locally integrable. Since the conditions of Theorem 11 are fulfilled, below, we discuss the conditions that guarantee the existence and uniqueness of the weak solution of the SDE (1.43) with exploding coefficient. The following Theorem is similar to Theorem 9.

Theorem 12. *Assume that the assumption 1 and the conditions of Theorem 9 hold. Then the solution X of the SDE* (1.43) *admits a weak solution, unique in the sense of the probability distribution and defined up to the explosion time.*

Proof. As the conditions of Theorem (11) are fulfilled and by applying the method of removal of drift (Theorem 8), we obtain

$$\hat{\sigma}(y) = \begin{cases} \sigma(\mathfrak{q}(y))H'(\mathfrak{q}(y)), & H(-\infty) < y < H(\infty), \\ 0 & \text{otherwise} \end{cases} \tag{1.58}$$

with H is given by

$$H(x) = K\int_c^x \exp\left(e^{-\frac{2}{\sigma^*}\int_{\bar{X}}^z g(y)dy}\right)dz \tag{1.59}$$

$$K = \exp(e^{-\frac{2}{\sigma^*}\int_{\bar{X}}^c g(y)dy}); \quad c > 0. \tag{1.60}$$

Note that $\bar{X} > x_e$ and c are two constants and therefore K is a constant. Because of the condition (2) and (1) of Theorem 11, we known that H' is bounded away from zero in finite intervals. For the rest of proof, we refer the reader to Theorem 9 in Karatzas and Shreve (2012). □

Remark 5. *We shown that if both drift anddiffusion coefficients are exploding, but the drift function explodes with magnitude higher than the one of the diffusion function, and then under the conditions of Theorem 12, the SDE 1.43 admits a weak solution which is unique in the sense of probability distribution and defined up to the exploding time.*

1.2.4 Examples of models generating drift burst (exploding models)

The model (1.43) in subsection 1.2.3 can be seen as the family of the exploding models. We recall that the coefficients of the model (1.43) depends on the unknown function g. In this subsection, we answer the question (2). Therefore, we discuss some examples of the function g. Note that each example of g corresponds to the exploding model. In several of these exploding models, we then verify if the conditions of Theorem (12) hold.

Example 3.

Suppose that a function g is of the form

$$g(x) = -\frac{a}{2(x-x_0)(k-a\ln(x-x_0))}, \quad x \in \mathbb{R}\setminus\{x_0, x_0+e^{k/a}\}, x_0 \in \mathbb{R}^+,$$

where $a, \sigma \in (0,\infty)$ *and* $k \in (0,\infty)$. *Substituting g in the expressions of drift anddiffusion coefficients in Eq.1.43 and we get:*

$$dX_t = -\frac{c(k-a\ln(X_t-x_0))^{\frac{1}{\sigma^*}-1}}{2(X_t-x_0)}dt + b(k-a\ln(X_t-x_0))^{\frac{1}{\sigma^*}}dW_t, \tag{1.61}$$

$$X_0 = \bar{X} > x_0 > 0, \quad t \in [0,T],$$

where $b = \frac{\sigma^*}{(k-a\log(\bar{X}-x_0)^{\frac{1}{\sigma^*}}}$ *and* $c = ab$. *We observe that both the drift anddiffusion coefficients of Eq.*(1.61) *explode at* $x = x_0$, *but the drift term converges to infinity faster than the diffusion term which is consistent with our theory discussed in 1). We also observe that for all* $x \in \mathbb{R}$ *the drift anddiffusion coefficients do not satisfy the regularity conditions such as the (local) Lipschitz and linear growth conditions since both functions are exploding at* $x = x_0$. *Let us now verify the conditions of Theorem* (11). *We prove that*

$$\frac{1}{(b(k-a\ln(X_t-x_0))^{\frac{1}{\sigma^*}})^2} \quad \textit{and} \quad \frac{a}{(b(X_t-x_0)(k-a\ln(X_t-x_0))^{1+\frac{1}{\sigma^*}})}$$

are locally integrable in $\mathbb{R}$. *At such point, choose* $\varepsilon > 0$ *so that* $-\infty < x+\varepsilon \leq x_0 < x+\varepsilon < \infty$ *and then we want to show*

$$\frac{1}{b^2}\int_{x-\varepsilon}^{x+\varepsilon} \frac{1}{(k-a\ln(z-x_0))^{\frac{2}{\sigma^*}}} dz \overset{?}{<} \infty. \tag{1.62}$$

By setting $y = \ln(x-x_0)$ *and then the integral* (1.62) *becomes*

$$\frac{1}{b^2}\int_{x-\varepsilon}^{x+\varepsilon} \frac{1}{(k-a\ln(z-x_0))^{\frac{2}{\sigma^*}}} dz = \frac{1}{(ab)^2}\int_{log(x-\varepsilon-x_0)}^{log(x+\varepsilon-x_0)} \frac{e^z}{(\frac{k}{a}-z)^{\frac{2}{\sigma^*}}} dz. \tag{1.63}$$

For sake of simplicity we set $\sigma^* = 1$ *and then by applying the integral by part, the second integral yields*

$$\begin{aligned}
\frac{1}{(ab)^2}\int_{log(x-\varepsilon-x_0)}^{log(x+\varepsilon-x_0)} \frac{e^z}{(\frac{k}{a}-z)^2} dz = {} & \frac{1}{(ab)^2}\frac{x+\varepsilon-x_0}{(\frac{k}{a}-log(x+\varepsilon-x_0))} - \frac{1}{(ab)^2}\frac{x-\varepsilon-x_0}{(\frac{k}{a}-log(x-\varepsilon-x_0))} \\
& + \frac{1}{(ab)^2}(x+\varepsilon-x_0)\log(\frac{k}{a}-\log(x+\varepsilon-x_0)) \\
& - \frac{1}{(ab)^2}(x-\varepsilon-x_0)\log(\frac{k}{a}-log(x-\varepsilon-x_0)) \\
& - (x+\varepsilon-x_0)\log(\frac{k}{a}-(x+\varepsilon-x_0)) + e^{\frac{k}{a}}Ei((x+\varepsilon-x_0)-\frac{k}{a}) \\
& + (x-\varepsilon-x_0)\log(\frac{k}{a}-(x-\varepsilon-x_0)) - e^{\frac{k}{a}}Ei((x-\varepsilon-x_0)-\frac{k}{a}),
\end{aligned} \tag{1.64}$$

where Ei is an exponential integral. [3] *The right side of* (1.64) *is finite, because whenever* $x-\varepsilon$ *converges to the exploding point* x_0 *all expressions at right side of* (1.64) *are finite. We can easily check it by setting* $a=1,k=2,\sigma^*=1,x_0=1,\bar{X}=2$. *Here the process explodes at* 1. *Let* $(1,2]\subset\mathbb{R}$ *be a compact set and then we write*

$$\frac{1}{b^2}\int_1^2 \frac{1}{(2-\ln(z-1))^2}dz=\frac{1}{4}(e^2Ei(-2)+\frac{1}{2})\approx\frac{1}{4}0.138671<\infty.$$

Thus, we conclude that $\frac{1}{(b(k-a\ln(X_t-x_0))^{\frac{1}{\sigma^*}})^2}$ *is locally integrable. To prove* $\frac{a}{(b(X_t-x_0)(k-a\ln(X_t-x_0))^{1+\frac{1}{\sigma^*}})}$ *is locally integrable, we adapt to the similar idea as the precedent proof and we also set* $\sigma^*=1$. *We have*

$$\begin{aligned}\frac{1}{ab^2}\int_{x-\varepsilon}^{x+\varepsilon}\frac{1}{(z-x_0)(\frac{k}{a}-\ln(z-x_0))^2}dz&=\frac{1}{ab^2}\frac{1}{(\frac{k}{a}-\ln(x+\varepsilon-x_0))}\\&-\frac{1}{ab^2}\frac{1}{(\frac{k}{a}-\ln(x-\varepsilon-x_0))}.\end{aligned}\tag{1.65}$$

The right side of (1.65) *is finite, because whenever* $x-\varepsilon$ *converges to the exploding point* x_0 *all expressions at right side of* (1.64) *are finite. We can check it as in the previous case by also setting* $a=1,k=2,\sigma^*=1,x_0=1,\bar{X}=2$. *Here the process still explodes at* 1. *Let* $[1,2]\subset\mathbb{R}$ *be a compact set and then we write*

$$\frac{1}{b^2}\int_1^2\frac{1}{(x-1)(2-\ln(z-1))^2}dz=\frac{1}{8}<\infty.$$

We conclude that $\frac{a}{(b(X_t-x_0)(k-a\ln(X_t-x_0))^{1+\frac{1}{\sigma^*}})}$ *is locally integrable. Since we shown that* σ^{-2} *and* $|\mu|\sigma^{-2}$ *of* (1.61) *are locally integrable and also* $\sigma^2>0$, *and then from the Theorem 12 we conclude that the solution of* (1.61) *admits a weak solution which is unique in the sense of probability and defined up to the explosion time.*

[3] An exponential integral is a function of the following form

$$E_i(x)=\int_{-x}^{\infty}\frac{e^{-t}}{t}dt=\int_{-\infty}^{x}\frac{e^t}{t}dt,\quad x>0,$$

see Abramowitz and Stegun (1965) and Temme (2011).

Example 4. *Suppose that a function g is of the form*

$$g(x) = \frac{a}{2(x-x_0)(k-\ln(x-x_0))}, \quad x \in \mathbb{R} \setminus \{x_0, x_0+e^k\},$$

where $a \in (-\infty, 0), k \in \mathbb{R}$. *Substituting g in the expressions of drift anddiffusion coefficients in Eq.*(1.43) *and we obtain:*

$$\begin{aligned} dX_t =& \frac{a\sigma^*[k-\ln(\bar{X}-x_0)]^{\frac{a}{\sigma^*}}}{2(X_t-x_0)(k-\ln(X_t-x_0))^{\frac{a}{\sigma^*}+1}}dt + \frac{\sigma^*(k-\ln(\bar{X}-x_0))^{\frac{a}{\sigma^*}}}{(k-\ln(X_t-x_0))^{\frac{a}{\sigma^*}}}dW_t, \\ & X_0 = \bar{X} > x_0 > 0, t \in [0,T], x_0 \in \mathbb{R}^+. \end{aligned} \tag{1.66}$$

We observe that both the drift anddiffusion coefficients of Eq.(1.66) *are exploding at* $x = x_0$, *but the drift term converges to infinity faster than the diffusion term (as consistent with our theory discussed in section 1). We also note for all* $x \in \mathbb{R}$ *the drift anddiffusion coefficients do not satisfy the regularity conditions such as the (local) Lipschitz and linear growth conditions since both functions are exploding at* $x = x_0$. *Note that the model* (1.66) *is similar to the model* (1.61). *In a similar way as in the Example 3, we can show the solution of* (1.66) *admits a weak solution which is unique in the sense of the probability distribution and defined up to explosion time.*

The following example differ from the previous ones:

Example 5. *Suppose that a function g is of the form*

$$g(x) = -\frac{a}{2(x-x_0)}, \quad x \in \mathbb{R} \setminus \{x_0\},$$

where $a \in (0, \infty)$. *Substituting g in the expressions of drift anddiffusion coefficients in Eq.* (1.43) *and we obtain:*

$$dX_t = -\frac{ab}{2(X_t-x_0)^{a/\sigma^*+1}}dt + \frac{b}{(X_t-x_0)^{a/\sigma^*}}dW_t, \quad X_0 = x \in \mathbb{R} \setminus \{x_0\}, t \in [0,\bar{T}]. \tag{1.67}$$

where $b = \sigma^*(\bar{X}-x_0)^{a/\sigma^*}, \bar{X} > x_0$. *Let us assume that* $a > \sigma^* > 0$ *and we observe that both the drift anddiffusion coefficients of Eq.*(1.66) *are exploding at* $x = x_0$, *but the drift term converges to infinity fast than the diffusion term (as consistent with our theory discussed in section 1). We also note for all* $x \in \mathbb{R}$ *the drift anddiffusion coefficients cannot satisfy the regularity conditions such as the (local) Lipschitz and linear growth conditions*

since both functions are exploding at $x = x_0$. Let now verify the conditions of Theorem (11). We prove that

$$\frac{(X_t - x_0)^{2a/\sigma^*}}{b^2} \quad \text{and} \quad \frac{a(X_t - x_0)^{a/\sigma^*-1}}{b},$$

are locally integrable in $\mathbb{R}$. At such point, choose $\varepsilon > 0$ so that $-\infty < x+\varepsilon < x_0 < x+\varepsilon < \infty$ and then write

$$\frac{1}{b^2}\int_{x-\varepsilon}^{x+\varepsilon}(z-x_0)^{2a/\sigma^*}dz = \frac{1}{b^2}\frac{1}{2a/\sigma^*+1}(x+\varepsilon-x_0)^{2a/\sigma^*+1} - \frac{1}{b^2}\frac{1}{2a/\sigma^*+1}(x-\varepsilon-x_0)^{2a/\sigma^*+1}. \tag{1.68}$$

The right parts of integral (1.68) are finite, because as $x-\varepsilon$ converges to the exploding point x_0 they are all finite since $2a/\sigma^*+1 > 0$ and thus $\frac{(X_t-x_0)^{2a/\sigma^*}}{b^2}$ is locally integrable. Let now show that $\frac{a(X_t-x_0)^{a/\sigma^*-1}}{b}$ is locally integrable in $\mathbb{R}$. In similar way as in the previous proof, we choose $\varepsilon > 0$ so that $-\infty < x+\varepsilon < x_0 < x+\varepsilon < \infty$ and also write

$$\frac{a}{b}\int_{x-\varepsilon}^{x+\varepsilon}(z-x_0)^{a/\sigma^*-1}dz = \frac{a}{b}\frac{1}{a/\sigma^*}(x+\varepsilon-x_0)^{a/\sigma^*} - \frac{a}{b}\frac{1}{a/\sigma^*}(x-\varepsilon-x_0)^{a/\sigma^*}. \tag{1.69}$$

We also observe that the right part of integral (1.69) are finite, because as $x-\varepsilon$ converges to the exploding point x_0 they are also finite since $2a/\sigma^* > 0$ and thus $\frac{(X_t-x_0)^{2a/\sigma^*}}{b^2}$ is locally integrable. Since we shown that σ^{-2} and $|\mu|\sigma^{-2}$ of (1.67) are locally integrable and $\sigma^2 > 0$, and from the Theorem 12 we conclude that the solution of (1.61) admits a weak solution which is unique in the sense of probability and defined up to the explosion time.

In the above discussion, we considered two different types of the function g. In the examples 1.61 and 1.66, we considered the case that the functions g have two singular points. In the example 1.67, we considered the case that g has one singular point. In all three cases, we saw that the drift anddiffusion coefficients explode at a unique point. These results imply that regardless of the type of g, the drift anddiffusion coefficients of the SDE (1.43) will explode at a unique point of $\mathbb{R}$. Therefore, we consider that the family of function g has the same properties as in the examples (3),(5) and (4):

Assumption 2. *Consider the SDE (1.43), then we have*

1. *A set of exploding point of the function g is might be made of single point or multiple points;*
2. *If g explodes at single point x_0, and then it is continuous on $\mathbb{R}\backslash\{x_0\}$. The integral $\int_\cdot^x g(u)du$ is finite for every $x \in \mathbb{R}\backslash\{x_0\}$ and $\int_\cdot^{x_0} g(u)du = +\infty$.*

Because of the assumption 2, the drift anddiffusion coefficients satisfy the following properties:

Proposition 1.

1. $\sigma \in C^3(\mathbb{R}\backslash\{x_0\})$,
2. $\sigma(x_0) = +\infty$;
3. $\mu \in C^1(\mathbb{R}\backslash\{x_0\})$;
4. $\mu(x_0^-) = \mu(x_0^+) = \pm\infty$.

Proof. The proof is straightforward and we omit all details. □

Remark 6. *Note that in the assumption (2) we consider the case $\int_0^{x_0} g(u)du = +\infty$. In case instead of that we have $\int_0^{x_0} g(u)du = -\infty$, one can consider $-g$ in place of g and act similarly. From Proposition (1), we can see that for every x in $\mathbb{R}$, the diffusion and drift coefficients(σ and μ) explode at a unique point x_0. We also observe that the drift anddiffusion coefficients and their corresponding derivatives are bounded in $x \in \mathbb{R} \setminus \{x_0\}$.*

In the following, we discuss how to estimate the parameters of the above parametric models discussed in examples 1.61, 1.66 and 1.67.

1.3 Parameter Estimation of SDEs with exploding coefficients

In the above section, the several parametric models of the SDE with exploding coefficients were discussed. In this section, we discuss how to estimate the parameters of these models. The family of the proposed model is of the form:

$$dX_t = \mu(X_t;\theta) + \sigma(X_t;\theta)dW_t; X_{t_0} = x_0 \tag{1.70}$$

with parameter vector θ. The drift and diffusion coefficients explode in the finite point with a drift explodes in the magnitude higher than that of a diffusion coefficient. Let us assume that a sample $X = \{X_{t_i}, i = 0, \cdots, n\}$ is available for analysis and the observations may not be equally spaced. We aim to estimate the set of the parameter θ of the model in Eq. (1.70) using the observation data. To estimate the set of the parameters of

the model in Eq. (1.70), we need to compute the maximum likelihood estimator. To compute the maximum likelihood estimator, we need that the transition density to be available. However, in many cases, the transition densities of the SDE are rarely available, so we need to rely on some existing methods to approximate the transition densities.

In literature, there exist many methods to approximate the transition density. For example, Kessler (1997) used a Gaussian distribution with the first and second moment computed from higher-order Ito-Taylor expansions to approximate the transition density. Shoji and Ozaki (1998) proposed method that allows us to linearize the SDE to obtain an approximation of Ornstein-Uhlenbeck process type (exact transition density of the Ornstein-Uhlenbeck process is known). Ait-Sahalia (1996) proposed a method to approximate a transition density based on Hermite function with coefficients computed from higher-order Ito-expansions. Other methods we refer to the works of Florens-Zmirou (1989) and Yoshida (1992). There are also many methods that have been proposed based on method of moments approaches, see for instance, Duffie and Singleton (1990), Gourieroux et al. (1993), Bibby and Sørensen (1995) and Gallant (1977). The methods mentioned above computed the expectation by using the simulation-based methods. We should point out that the methods mentioned above do not approximate the transition density of (1.70) directly, but they do approximate the transition density of the discretized version of the model (1.70).

Maximum-likelihood estimation procedure

The time homogeneous stochastic differential equation is

$$dX_t = \mu(X_t;\theta) + \sigma(X_t;\theta)dW_t; X_{t_0} = x_0. \tag{1.71}$$

Suppose that $t_i = \frac{iT}{n}; \quad i = 0,\ldots,n$ is a discretization of the time interval $[0,T]$ with discretization step Δ is setted to be equal to $\Delta = \frac{T}{n}$. We set $\Delta W_i = W_{t_{i+1}} - W_{t_i}, \quad i = 0,\ldots,n-1$ where ΔW_i the increment of the Weiner process with independent and identically distributed normal random variables with expected value zero and variance Δ. Suppose that $X = \{X_{t_i} = X_i; i = 1,\cdots,n\}$ is a the solution of process (1.71), that is

$$X_{i+1} = X_i + \int_{t_i}^{t_{i+1}} \mu(X_s)ds + \int_{t_i}^{t_{i+1}} \sigma(X_s)dW_s \tag{1.72}$$

Since coefficients of an SDE (1.71) do not depend on time t, and we can assume that the transition density of (1.71) depends only on the discretization step Δ and the stochastic processes x, y. The transition density of (1.71) can be expressed in the following form $p(\Delta, x, y)$. Since the transition density of the family of exploding process (1.71) is not available, we cannot in principle obtain the Likelihood. Nevertheless, we can still approximate the minimize negative log-likelihood of the unknown transition densities based on some existing methods such as the Euler-Maruyama method (Preisler et al. 2013 and Brillinger 2010) and the method proposed by Shoji and Ozaki (1998).

Euler-Maruyama method

Suppose that $\widehat{X} = \{\widehat{X}_t, t \in [0,T]\}$ is approximated solution of X by Euler -Maruyama method method and satisfying the following recursive equation

$$\widehat{X}_{i+1} = \widehat{X}_i + \mu(\widehat{X}_i)\Delta + \sigma(\widehat{X}_i)\Delta W_i, \tag{1.73}$$

for $i = 0, \ldots, n-1$ with initial value $\widehat{X}_0 = X_0$. We observe that the increments $\widehat{X}_{i+1} - \widehat{X}_i$ are then independent Gaussian random variable with mean $\mathbb{E}_x[\widehat{X}_{i+1} - \widehat{X}_i] = \mu(\widehat{X}_i)\Delta$, and the variance $V_x(\widehat{X}_{i+1} - \widehat{X}_i) = \sigma(\widehat{X}_i)^2\Delta$. Let $p_\theta(\cdot)$ be a transition density. The approximate transition density of the process $\widehat{X}$ is given by:

$$p_\theta(\Delta, y|x) = \frac{1}{\sqrt{2\pi\sigma(x)^2\Delta}} \exp\left(- \frac{(y - x - \mu(x)\Delta)^2}{2\sigma(x)^2\Delta} \right), \tag{1.74}$$

and then the log-likelihood is given by

$$L_n(\theta) = \sum_{i=1}^{n} \log p_\theta(\Delta, \widehat{X}_{i+1}, \widehat{X}_i) = -\frac{1}{2} \sum_{i=1}^{n} \left(\frac{(\widehat{X}_{i+1} - \widehat{X}_i - \mu(\widehat{X}_i)\Delta)^2}{\sigma^2(\widehat{X}_i)\Delta} + \log(2\pi\sigma(\widehat{X}_i)^2\Delta) \right). \tag{1.75}$$

The equation (1.79) is also called locally Gaussian approximation and θ represents $r-$dimension of parameters of interest. In order to estimate the parameters vector θ, we have to minimize the negative log-likelihood (1.79) with respect to θ, e.i.

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} (-L_n(\theta)), \tag{1.76}$$

where $\Theta \subset \mathbb{R}$ is a set of unknown parameters. It should be noted that the maximum likelihood estimator obtained with the Euler method can be consistent under certain assumptions on the number of observation n and the discretization step Δ_n, which is assumed to converge to zero when the number of observation n increases. However, there is no guarantee that the estimator will remain unbiased when estimating the parameters using the flash crash data. It is, therefore, reasonable to consider another method which different from the Euler-Maruyama method. Thus, we consider the higher-order scheme proposed by Shoji and Ozaki (1998), which allows us to locally linearize the SDE to obtain the Ornstein-Uhlenbeck process of type.

Ozaki method

The Ozaki linearization method proposed by Shoji and Ozaki (1998) and Ozaki (1992), consists to linearize the SDE (1.70) in order to obtain an Ornstein-Uhlenbeck process. From the model (1.70), we can see that the diffusion function of (1.70) is not constant, however, to use Ozaki method we require that a diffusion coefficient of (1.70) needs to be constant. To obtain a constant diffusion of (1.70), we change the state variable from X to Y as:

$$Y_t = \Phi(X_t) = \int_t^X \frac{1}{\sigma(z;\theta)} dz \tag{1.77}$$

with Φ is a smooth function. Ito's lemma with respect to the SDE (1.70) yields

$$\begin{aligned} dY_t &= \left(\frac{\mu(X_t;\theta)}{\sigma(X_t;\theta)} - \frac{1}{2}\sigma'(X_t;\theta) \right) dt + dW_t \\ &= \left(\frac{\mu(\Phi^{-1}(Y_t);\theta)}{\sigma(\Phi^{-1}(Y_t);\theta)} - \frac{1}{2}\sigma'(\Phi^{-1}(Y_t);\theta) \right) dt + dW_t \\ &= \widehat{\mu}(Y_t;\theta)dt + dW_t, \end{aligned} \tag{1.78}$$

where σ' is a derivative and Φ^{-1} is the inverse of Φ. Since the equation (1.78) has a constant diffusion coefficient, we can apply the local linearization method of Shoji and Ozaki (1998) to the discretization process of (1.78). By using the Markov propriety of discretized process, we then can approximate the log-likelihood of the transition

density of (1.78) as:

$$L_n^o(\theta) = \sum_{i=1}^{n} \log p_\theta(\Delta, Y_{i+1}, Y_i) = -\frac{1}{2}\sum_{i=2}^{n}\left(\frac{(Y_i - E_{i-1})^2}{V_{i-1}} + \log(2\pi V_{i-1})\right), \tag{1.79}$$

where

$$\begin{aligned}
E_{i-1} &= Y_{i-1} + \frac{\widehat{\mu}(Y_{i-1};\theta)}{L_{i-1}}\Big(\exp(L_{i-1}\Delta) - 1\Big) + \frac{M_{i-1}}{L_{i-1}^2}\Big(\exp(L_{i-1}\Delta) - 1 - L_{i-1}\Delta\Big),\\
V_{i-1} &= \left(\frac{\exp(L_{i-1}\Delta) - 1}{2L_{i-1}}\right)\sigma^2,\\
L_{i-1} &= \widehat{\mu}'(Y_{i-1};\theta),\\
M_{i-1} &= \frac{\sigma}{2}\widehat{\mu}''(Y_{i-1};\theta),\\
\sigma &= 1.
\end{aligned}$$

The Eq. (1.80) allows us to approximate the transition density of Y. However, we want to approximate the transition density of X. To do so, we rely on the transformation of density rule. Using the Jacobian formula, we approximate the log-likelihood of the transition density of (1.70) as follows:

$$\sum_{i=1}^{n} \log p_\theta(\Delta, X_{i+1}, X_i) = \sum_{i=1}^{n} \log p_\theta(\Delta, Y_{i+1}, Y_i) + \sum_{i=1}^{n} |\Phi'(X_i)|. \tag{1.80}$$

The equation (1.79) is an approximation of transition density of X. Transforming the exploding model to obtain the model with constant volatility as discussed above, may lead to some complication because the inverse function $\Phi(.)^{-1}$ in Eq. (1.77)) may not exists in the close form. To address these issues, we follow the suggestion of Nowman (1997). That means when applying the local linearization approach to the model with non-constant diffusion coefficient, we only need to treat the non-constant diffusion coefficient like it is constant. Instead, using the Eq. (1.77) to transform the model (1.70) to the model (1.78). Thus, the log-likelihood of the transition density of (1.70) approximates as follows:

$$L_n^o(\theta) = \sum_{i=1}^{n} \log p_\theta(\Delta, X_{i+1}, X_i) = -\frac{1}{2}\sum_{i=2}^{n}\left(\frac{(X_i - E_{i-1})^2}{V_{i-1}} + \log(2\pi V_{i-1})\right), \tag{1.81}$$

where

$$
\begin{aligned}
E_{i-1} &= X_{i-1} + \frac{\mu(X_{i-1};\theta)}{L_{i-1}}\Big(\exp(L_{i-1}\Delta) - 1\Big) + \frac{M_{i-1}}{L_{i-1}^2}\Big(\exp(L_{i-1}\Delta) - 1 - L_{i-1}\Delta\Big),\\
V_{i-1} &= \left(\frac{\exp(L_{i-1}\Delta) - 1}{2L_{i-1}}\right)\sigma^2(X_{i-1};\theta),\\
L_{i-1} &= \mu'(Y_{i-1};\theta),\\
M_{i-1} &= \frac{\sigma(X_{i-1};\theta)}{2}\mu''(X_{i-1};\theta),
\end{aligned}
$$

To estimate the parameters θ, we minimize the negative likelihood in Eq. (1.80) or (1.81) with respect to θ as in the case of Euler-Maruyana method.

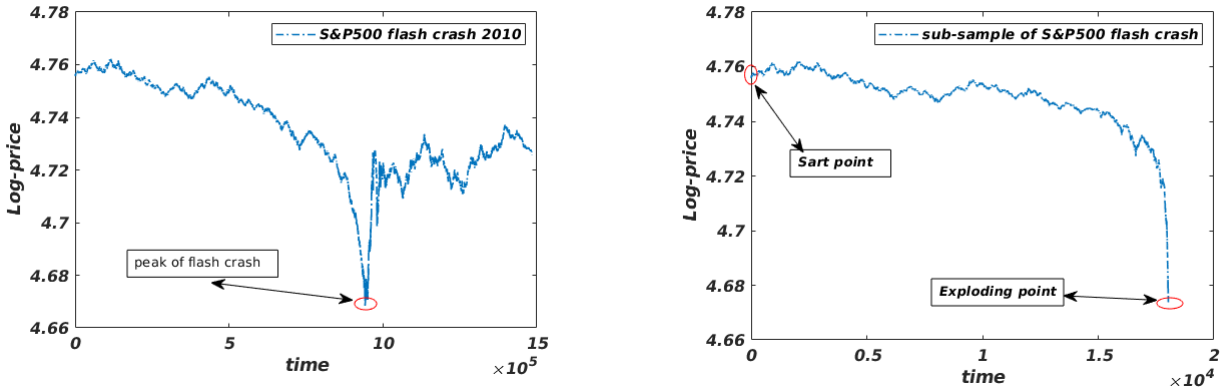
In Durham and Gallant (2002), a study comparing many different discrete maximum likelihood (DLM) methods was conducted. They found that the local linearization method is among the most accurate DLM estimator. They also suggested that the most of DLM methods in which their implementation may not require the used of transformation discussed above, show significant improvement if we transform the SDE into the one with constant drift using the transformation in Eq. (1.78). It is maybe because the transition distribution of X is probably far from the normal transition, while that of the new process Y is perhaps closer to the normal distribution (see Jeisman 2006). We should be noted that the advantages of using DLM methods are that they are simple to implement and that they do not require any restriction on the coefficients of the SDE, which is very important, especially in our case where the drift and diffusion coefficients explode in the finite point.

1.3.1 The Data

In this subsection, we consider the E-mini S&P 500 flash crash data of May 6, 2010 (SPY). Figure 2.4(a) plots the log-price of SPY flash crash data. We observe that the peak of the log-price of the SPY flash crash is 4.6738 as indicated on red point. Based on the theory of Section 1.2, we consider the point 4.678 of SPY flash crash to be the exploding point of SPY flash crash. To correctly estimate the parameters of exploding models discussed in Section 1.2 on the SPY flash crash data, we have to be consistent with the theory of Section 1.2. Therefore, we cannot estimate the parameters of exploding models on the full sample of SPY flash crash data. Thus, we need to fit the exploding parametric models on the sub-sample of SPY flash crash. Figure 2.4(b) shows the

trajectory of the sub-sample of SPY flash crash data. In Table 1.1, we summarize the statistics of the sub-sample of E-mini S&P 500 flash crash of May 6, 2010. From this table, we see that the mean is equal to 4.7491 and the standard deviation is 0.0096. The total number of observations of sub-sample of SPY flash crash data is equal $18,026$, and the price before the explosion is equal to 4.791. We set the time step $\Delta = \frac{1}{252\times 23400}$ so that the estimated parameters expressed in annualized form.

Figure 1.1 Left panel shows log-price of S&P 500 flash crash on $06/05/2010$ and a right panel is his sub-sample.



(a) (b)

Table 1.1 Summary Statistics of E-mini S&P 500 flash crash, May 2010. The N denotes the total number of observations of sub-sample of SPY flash crash data, μ the mean and $\bar{\sigma}$ the standard deviation. The $\bar{X}$ denotes a price before the flash crash starts and x_0 is the exploding price or peak of the flash crash.

SPY flash crash				
N	μ	$\bar{\sigma}$	$\bar{X}$	x_0
18026	4.7491	0.0096	4.7558	4.6738

1.3.2 Empirical application

The main objective of this section is to estimate the parameters of the models introduced in Section 1.2. In Table (3.4) below, we present the exploding parametric models under consideration. We fit each parametric model to the sub-sample of S&P 500 flash crash. We estimate these parameters by the maximum likelihood estimation based on the solution of the parametric models $1,2$ and 2 discretized by the Euler-Maruyama and local linearization methods. We use Akaike's Information criteria (AIC, Akaike 1974a and Bayesian Information

Criteria (BIC,Schwarz et al. 1978)) to select the best model among the three. The model with the lowest values of AIC or BIC will be considered as the best.

Table 1.2 Summary of the exploding models introduced in section 1.2.4.

Exploding Model
Model 1
$$dX_t = -\frac{a\sigma^*(k-a\ln(X_t-x_0))^{\frac{1}{\sigma^*}-1}}{2(X_t-x_0)(k-a\log(\bar{X}-x_0)^{\frac{1}{\sigma^*}}}dt + \frac{\sigma^*(k-a\ln(X_t-x_0))^{\frac{1}{\sigma^*}}}{(k-a\log(\bar{X}-x_0)^{\frac{1}{\sigma^*}}}dW_t,$$
Model 2
$$dX_t = \frac{a\sigma^*[k-\ln(\bar{X}-x_0)]^{\frac{a}{\sigma^*}}}{2(X_t-x_0)(k-\ln(X_t-x_0))^{\frac{a}{\sigma^*}+1}}dt + \frac{\sigma^*(k-\ln(\bar{X}-x_0))^{\frac{a}{\sigma^*}}}{(k-\ln(X_t-x_0))^{\frac{a}{\sigma^*}}}dW_t$$
Model 3
$$dX_t = -\frac{ab}{2(X_t-x_0)^{a/\sigma^*+1}}dt + \frac{b}{(X_t-x_0)^{a/\sigma^*}}dW_t, \quad b=\sigma^*(\bar{X}-x_0)^{a/\sigma^*}.$$

In Tables 1.3 and 1.4, we report the parameters of the exploding models estimated using the sub-sample of SPY flash crash price. In Table 1.3, we present the estimated parameters of the exploding models whose densities distributions approximated based on the Euler-Maruyama method. And in 1.4, we display the estimated parameters of the exploding models whose densities distributions approximated based on the local linearization method. From the tables 1.3 and 3.1, we can see that the estimated parameters of the models 1 and 2 obtained with the Euler approximation are not different from those obtained with the local linearization method. Furthermore, the estimated parameters of the models 1 and 2 obtained with the Euler approximation are similar to those obtained with the local linearization approximation. On the other hand, the estimated parameters of model 3 obtained with the Euler-Maruyama method are slightly different from those obtained with local linearization method. The estimated parameter a obtained with the Euler approximation is slightly different from that obtained

with the local linearization approximation, and the estimated parameter σ^* obtained with the Euler method is identical to that obtained with the local linearization method. From the Tables 1.3 and 3.1, we can also see the model 3 fits the data better compared to the models 1 and 2 according to the BIC and AIC results. These results suggest that the model 3 performs better than the model 2 and 1, although the model 1 and 2 have more parameters than the models 3.

Table 1.3 Table (3.3) reports the MLE for the parameters of three exploding models estimated using the E-mini S&P 500 flash crash of May, 2010. We calculate the parameters using the Euler-Maruyama approximation and between parentheses are standard errors.

Maximum Likelihood estimation based on the Euler-Maruyama method						
Model#	Estimated Parameters			Log-likelihood	AIC	BIC
1	k	a	σ^*			
	5.583505	0.313882	0.278899	137059.07	-274114.14	-274098.54
	(0.264936)	(0.015503)	(0.002571)			
2	1.856764	-0.365213	0.265888	137563.54	-275123.07	-275107.47
	(0.000000)	(0.011218)	(0.001596)			
	a	σ^*				
3	0.306066	0.229874		139077.46	-278148.93	-278125.53
	(0.001375)	(0.001375)				

We conclude that the Euler method and the local linearization methods produce similar results when one estimates the parameters of exploding processes.

Table 1.4 Table (1.4) reports the MLE for the parameters of three exploding models estimated using the E-mini S&P 500 flash crash of May, 2010. We calculate the parameters using the local linearization approximation and between parentheses are standard errors.

Discrete Maximum Likelihood estimation based on the Ozaki & Shoji method						
Model#	Estimated Parameters			Log-likelihood	AIC	BIC
1	k	a	σ^*			
	5.583505	0.313882	0.278899	137059.19	-274663.4	-274098.64
	(0.000000)	(0.009862)	(0.001735)			
2	1.856764	-0.365213	0.265888	137563.58	-275123.17	-275107.57
	(0.000000)	(0.010899)	(0.001685)			
	a	σ^*				
3	0.305605	0.229919		139077.04	-278148.088	-278124.69
	(0.005965)	(0.001373)				

In the following we propose the method on how to simulate the models discussed in the Examples 1.61, 1.66 and 1.67.

1.4 Euler discretization of exploding process

1.4.1 Introduction

In Section 1.1, we discussed some example of stochastic differential equations (SDEs) with exploding coefficients (exploding models). As highlighted in Section 1.1 that the exploding models play a significant role in the field of financial mathematics. For example, the exploding models are used to simulate the trajectory of stochastic phenomena such as the flash crash prices (see Christensen et al. 2016 and Section 1.1). Unfortunately, in many cases, analytical solutions of SDEs are not available, and therefore the development of efficient numerical methods to simulate the trajectories of these SDEs is an important research topic.

In the literature, there exist several numerical methods for simulating the trajectories of SDEs. Here are a few references: Pardoux and Talay (1985), Talay (1990), Saito and Mitsui (1993), Bally and Talay (1996), Platen (1999), Higham (2001), Tocino and Ardanuy (2002), Rubenthaler (2003), Carletti et al. (2004), Sauer (2012), Bayram et al. (2018). The quoted developed numerical methods concerned continuous or jump processes. Our objective is to simulate the exploding trajectories like that observed, for example, during the flash crash of May 2010. Therefore, we cannot rely on the aforementioned numerical method to simulate more accurately the trajectories of exploding processes (SDEs with exploding coefficients). We propose a dynamic Euler discretization method in which we can choose the discretization step at each step of the recursion to obtain the desired root mean square. The latter is approximated using a toy model for which the analytical solution is available.

1.4.2 Euler-Maruyama method

Let $(\Omega, \mathscr{P}, (\mathscr{F}_t)_{t\geq 0}, \mathscr{F})$ be filtered probability space satisfying the usual conditions and $\{W_t\}_{t\in[0,T]}$ a standard Brownian. We return to the SDE in Eq.(1.43) of Section 1.1:

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dW_t, \quad Y_0 = Y_{t_0} \tag{1.82}$$

defined over the time interval $[0,T]$. As stated above, the diffusion (σ) and drift (μ) coefficients diverge to infinity in the finite point of $\mathbb{R}$. It is well-known that the analytical solutions of SDE in Eq (1.82) are rarely available, so we are required to use numerical methods such as Euler's Maruyama method to approximate a

solution process of Y. The Euler method named after Leonhard Euler and introduced around $1768-1870$, it is one of the simplest numerical methods and was first applied by Maruyama (1955) to stochastic differential equations.

Consider a following sequence of times:

$$t_0 = 0 < t_1 < t_2 < t_3 \cdots < t_{n+1} = T,$$

and the corresponding discretization step $\Delta t_i = t_{+1} - t_i$ with time $t_i = \frac{iT}{n}$. Let $\Delta W_i = W_{t_{i+1}} - W_{t_i}, \quad i = 0, \ldots, n-1$ be the increment of the Weiner process with independent and identically normally distributed random variable with mean zero and variance Δ. The one-dimensional Euler–Maruyama (EM) method for the Itô process 1.82 is the following recursive equation:

$$Y_{i+1}^{Eul} = Y_i^{Eul} + \mu(Y_i^{Eul})\Delta t_i + \sigma(Y_i^{Eul})\Delta W_i, \tag{1.83}$$

where Y_i^{Eul} is the numerical solution of 1.82 at time t_i. Approximate solutions in Eq.1.83 generally uses to simulate the accurate continuous trajectory of the solution process of SDEs. However, as noted above, the SDE trajectory does not always show a smooth behaviour. For example, the SDE trajectory in Eq. 1.82 explodes at a single point in $\mathbb{R}$. In this situation, the Euler Maruyama method may simulate more accurately the SDE trajectory in Eq. 1 only up to a point before the explosion, because in the exploding region, the discretization step must be taken very small. Keeping the discretization step fixed may cost the accuracy of Euler Maruyama method when simulating an exploding process. It is, therefore, reasonable that one needs to choose the discretization step in function of the variation of the trajectory of SDE in every step of the simulation. Our objective is to propose the algorithm that allows us to choose the discretization step in every step of simulation (dynamic discretization step) in such a way of hoping that the approximate solution and analytical solution are identical even in the exploding region.

1.4.3 Optimal Euler discretization of exploding processes

In this section, we discuss how to build the algorithm that allows us to simulate more accurately an exploding model. As mentioned above, the analytical solutions of exploding models are not available. So we first consider

a toy of exploding models in which the analytical available. We then propose a dynamic Euler algorithm of a toy model in which one can choose the discretization step at every step of recursion to obtain the desired root mean square error. Lastly, we apply the algorithm to the family of exploding models which are different from the toy model.

1.4.4 Dynamic Euler algorithm of a toy model

We built a dynamic Euler algorithm of a toy model which helps us to choose the discretization step dynamically to obtain the desired root mean square error. The algorithm will allow one to simulate more accurately a trajectory of a toy mode of exploding models.

Toy Model

The toy model is

$$\begin{aligned} dY_t &= c_\mu\Big(\frac{1}{(t_{end}-t)^\alpha}-1\Big)dt + c_\sigma\Big(\frac{1}{(t_{end}-t)^\beta}-1\Big)dW_t \\ &= \mu(t)dt+\sigma(t)dW_t, \end{aligned} \tag{1.84}$$

where $\beta \in (0,0.5)$, $\alpha \in (0.5,1)$, c_μ and c_σ are constants. The term W is a one-dimension standard Brownian motion and t_{end} is the exploding time. Note that the drift and diffusion coefficients of the toy model are a function of time instead of stochastic process as in Eq. (1.61). We also observe that the drift and diffusion coefficients of the toy model are deterministic, and we can, therefore, compute its analytical solution.

Let Y denotes the analytical solution of the toy model in Eq- (1.84), and it is given by:

$$\begin{aligned} \Delta Y_i = Y_{t_{i+1}} - Y_{t_i} &= -c_\mu\Delta t_i + c_\mu\int_{t_i}^{t_{i+1}}\frac{1}{(t_{end}-s)^\alpha}ds + c_\sigma\int_{t_i}^{t_{i+1}}\Big[\frac{1}{(t_{end}-s)^\beta}-1\Big]dW_s \\ &= -c_\mu\Delta t_i - \frac{c_\mu}{1-\alpha}(t_{end}-t_{i+1})^{-\alpha+1} + \frac{c_\mu}{1-\alpha}(t_{end}-t_i)^{-\alpha+1} + c_\sigma\int_{t_i}^{t_{i+1}}\Big[\frac{1}{(t_{end}-s)^\beta}-1\Big]dW_s. \end{aligned} \tag{1.85}$$

Suppose that Y^{Eul} is an approximate solution of a toy model by Euler-Maruyama method defines as follows:

$$\Delta Y_i^{Eul} = Y_{t_{i+1}}^{Eul} - Y_i^{Eul} = c_\mu \left[\frac{1}{(t_{end}-t_i)^\alpha} - 1\right]\Delta t_i + c_\sigma \left[\frac{1}{(t_{end}-t_i)^\beta} - 1\right]\Delta W_i. \tag{1.86}$$

To verify how accurate the Euler-Maruyama method approximates the analytical solution of a toy model, we use the root means square (RMSE). The RMSE is the expectation of the square root of the difference between the analytical solution and the approximate one, that is:

$$RMSE(\%) = (\mathbb{E}[(\Delta Y_i^{Eul} - \Delta Y_i)^2])^{1/2} \times 100, \quad \Delta y_i = y_{i+1} - y_i, \tag{1.87}$$

or

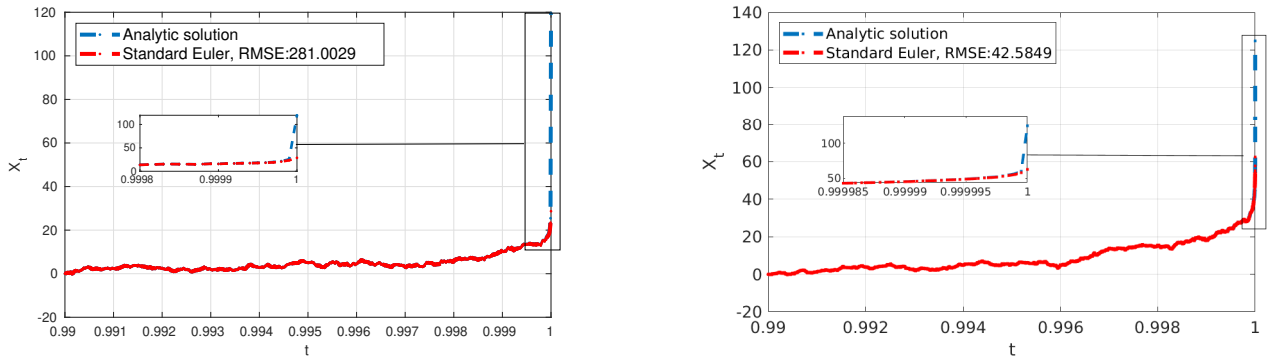
$$RMSE(\%) = \sqrt{\frac{\sum_{i=0}^{n}(\Delta \hat{Y}_i - \Delta Y_i)^2}{n}} \times 100. \tag{1.88}$$

The small value of RMSE will imply that the Euler-Maruyama method approximates more accurately the analytical solution. On the other hand, we also know that the accuracy of Euler-Maruyama methods depends critically on the size of the discretization step. However, the fact that the trajectory of a toy model varies considerably around the exploding time t_{end}, and makes it challenging to choose the correct discretization. It is, therefore, reasonable to consider the discretization step with different size of n. Thus, we consider the discretization step with two different sizes, e.i. $\Delta t_i = 5 \times 10^{-5}$ and 5×10^{-7}. We set$\sigma = 0.1, \alpha = 0.999, c_\sigma = 100, c_\mu = 5$ and $t_{end} = 1$.We simulate the analytical and approximate solutions using the two different Δt_i. For each simulation, the number of iteration stops once the process hits the exploding time.

In Figures 1.2(a) and 1.2(b), we plot the trajectories of the toy model simulate using the analytic formula (red-line) and Maruyama method (blue-line) with different discretization step, $\Delta t_i = 1 \times 10^{-5}$ and $\Delta t_i = 5 \times 10^{-7}$. Let us start with Figure 1.2(a). Figure 1.2(a) shows that the trajectory simulates with the Euler-Maruyama method distinct significantly from the one simulates with the analytical solution around the exploding time (see also sub-Figure 1.2(a)). This difference is further confirmed by looking at the value of RMSE as the latter is almost 281%, which is very large. On the other hand, the trajectory of a toy model simulates using the Euler Maruyama method with a small discretization step remains distinct to the one generated with the analytic formula

around the exploding time as we can see from Figure 1.2(b). The results may suggest the Euler-Maruyama method seems to improve its accuracy when the discretization step is taken very small. However, the Euler Maruyama method still not able to simulate more accurately the trajectories of a toy model around the exploding region regardless of our choice of the discretization step. The main question arises here is: how to choose the discretization step Δt_i in such a way of hoping that the Euler approximate solution and the analytical solution are virtually identical around the exploding point? In the following, we respond to the question.

Figure 1.2 The analytical solution (blue dot line) and Euler-Maruyama approximate solution (red dot line) as well the RMSE in percentage. Left panel 1.2(a): compares the trajectoriesof a toy model simulate with the analytical and Euler-Maruyama approximate solutions for $\Delta t_i = 5 \times 10^{-5}$. Left panel(1.2(b)): compares the trajectoriesof a toy model simulate with the analytical and Euler-Maruyama approximate solutions for $\Delta t_i = 5 \times 10^{-7}$. Here the constants $\sigma = 0.1, \alpha = 0.999, c_\sigma = 100, c_\mu = 5$.



(a) (b)

Algorithm to determine the discretization step

In this section, we answer the above question. We will propose an algorithm which allows us to select the discretization dynamically step at each step of the recursion to obtain the target root means square error (RMSE).

The RMSE in Eq. (1.87) can also be written in the following form:

$$\mathbb{E}[(\Delta Y_i^{Eul} - \Delta Y_i)^2] = \mathbb{E}[(\Delta Y_i)^2] - 2\mathbb{E}[\Delta Y_i \Delta Y_i^{Eul}] + \mathbb{E}[(\Delta Y_i^{Eul})^2], \tag{1.89}$$

where Y^{Eul} is given in Eq. (1.86) and Y in Eq. (1.85). We can compute each term in Eq. (1.89) as follows:

$$\mathbb{E}[(\Delta Y_i)^2] = c_\mu^2\Big(-\Delta t_i + \int_{t_i}^{t_{i+1}} \frac{1}{(t_{end}-s)^\alpha} ds\Big)^2 + c_\sigma^2\Big(\int_{t_i}^{t_{i+1}} \Big[\frac{1}{(t_{end}-s)^{2\beta}} - \frac{2}{(t_{end}-s)^\beta} + 1\Big] ds\Big) \tag{1.90}$$

$$\mathbb{E}[(\Delta Y_i^{Eul})^2] = c_\mu^2\Big[\frac{1}{(t_{end}-t_i)^\alpha} - 1\Big]^2 \Delta t_i^2 + c_\sigma^2\Big[\frac{1}{(t_{end}-t_i)^\beta} - 1\Big]^2 \Delta t_i \tag{1.91}$$

$$\begin{aligned}\mathbb{E}[(\Delta Y_i \Delta Y_i^{Eul}] =& c_\mu^2\Big[\frac{1}{(t_{end}-t_i)^\alpha} - 1\Big]\Delta t_i \int_{t_i}^{t_{i+1}} \Big[\frac{1}{(t_{end}-s)^\alpha} - 1\Big] ds \\ &+ c_\sigma^2\Big[\frac{1}{(t_{end}-t_i)^\beta} - 1\Big]\mathbb{E}[\int_{t_i}^{t_{i+1}} dW_s \int_{t_i}^{t_{i+1}} \Big[\frac{1}{(t_{end}-s)^\beta} - 1\Big] dW s] \\ =& c_\mu^2\Big[\frac{1}{(t_{end}-t_i)^\alpha} - 1\Big]\Delta t_i \int_{t_i}^{t_{i+1}} \Big[\frac{1}{(t_{end}-s)^\alpha} - 1\Big] ds + c_\sigma^2\Big[\frac{1}{(t_{end}-t)^\beta} - 1\Big] \int_{t_i}^{t_{i+1}} \Big[\frac{1}{(t_{end}-s)^\beta} - 1\Big] ds.\end{aligned} \tag{1.92}$$

Substituting the expressions (1.90), (1.91) and (1.92) in Eq. (1.89), and we get

$$\begin{aligned}\mathbb{E}[(\Delta Y_i^{Eul} - \Delta Y_i)^2] =& c_\mu^2\Big(\frac{1}{(t_{end}-t_i)^\alpha}\Delta t_i - \frac{1}{1-\alpha}(t_{end}-t_{i+1})^{-\alpha+1} + \frac{1}{1-\alpha}(t_{end}-t_i)^{-\alpha+1}\Big)^2 \\ &+ c_\sigma^2\Big[\frac{1}{(t_{end}-t_i)^\beta} - 1\Big]^2 \Delta t_i - c_\sigma^2 \frac{1}{1-2\beta}(t_{end}-t_{i+1})^{-2\beta+1} + c_\sigma^2 \frac{1}{1-2\beta}(t_{end}-t_i)^{-2\beta+1} \\ &+ c_\sigma^2 \frac{2}{1-\beta}(t_{end}-t_{i+1})^{-\beta+1} - c_\sigma^2 \frac{2}{1-\beta}(t_{end}-t_i)^{-\beta+1} + c_\sigma^2 \Delta t_i \\ &- 2c_\sigma^2\Big[\frac{1}{(t_{end}-t)^\beta} - 1\Big]\Big[-\frac{1}{1-\beta}(t_{end}-t_{i+1})^{-\beta+1} + \frac{1}{1-\beta}(t_{end}-t_i)^{-\beta+1} - \Delta t_i\Big].\end{aligned} \tag{1.93}$$

Now let RMSE* be our target RMSE. The difference between expected value in Eq. (1.93) and RMSE*, denotes by $D(\cdot)$ and it is given by:

$$\begin{aligned}D(\Delta t_i) =& c_\mu^2\Big(\frac{1}{(t_{end}-t_i)^\alpha}\Delta t_i - \frac{1}{1-\alpha}(t_{end}-t_{i+1})^{-\alpha+1} + \frac{1}{1-\alpha}(t_{end}-t_i)^{-\alpha+1}\Big)^2 \\ &+ c_\sigma^2\Big[\frac{1}{(t_{end}-t_i)^\beta} - 1\Big]^2 \Delta t_i - c_\sigma^2 \frac{1}{1-2\beta}(t_{end}-t_{i+1})^{-2\beta+1} + c_\sigma^2 \frac{1}{1-2\beta}(t_{end}-t_i)^{-2\beta+1} \\ &+ c_\sigma^2 \frac{2}{1-\beta}(t_{end}-t_{i+1})^{-\beta+1} - c_\sigma^2 \frac{2}{1-\beta}(t_{end}-t_i)^{-\beta+1} + c_\sigma^2 \Delta t_i \\ &- 2c_\sigma^2\Big[\frac{1}{(t_{end}-t)^\beta} - 1\Big]\Big[-\frac{1}{1-\beta}(t_{end}-t_{i+1})^{-\beta+1} + \frac{1}{1-\beta}(t_{end}-t_i)^{-\beta+1} - \Delta t_i\Big] - \text{RMSE}^*\end{aligned} \tag{1.94}$$

where , $i = 1,2,\cdots,n-1$. To determine the discretization step $\hat{\Delta}_j$ at each step of the iteration j, we need to minimize the expression (1.94) as follows:

$$\hat{\Delta}_j = \arg\min_{\Delta_j} D(\Delta_j), \quad j = 1,2,\ldots \tag{1.95}$$

where $\Delta_j = (t_{end} - t\,j)/2$. Using the estimated discretization step $\hat{\Delta}^j$, we define the dynamic Euler approximation to Y_T by $\widetilde{Y}_0 = Y_0$ as:

$$\widetilde{Y}_{j+1} = \widetilde{Y}_j + \mu(t_j)\hat{\Delta}t_j + \sigma(t_j)\hat{\Delta}W_j, \quad j = 0,1,2,\ldots, \tag{1.96}$$

For example, to choose a discretization step$\hat{\Delta}_0$ between $\bar{Y}_{t_0}$ and $\bar{Y}_{t_1}$, we need to minimize the equation (1.95) with respect to $(t_{end} - t_0)/2$. Then, we will keep estimating $\hat{\Delta}$ at every iteration until the process stops once it hits the exploding time t_{end}. Below, we now present the algorithm to implement the dynamic Euler approximation of a toy model.

Algorithm 1 Algorithm to implement Dynamic Euler of a toy model

---

procedure ITERATION(t_{start}, t_{end}, max-steps) ▷ Initialize all input parameters

  $t_{start} \leftarrow$ Starting point

  $t_{end} \leftarrow$ Exploding point ,

  $j \leftarrow 0$

  while $(t_{start} < t_{end})$ **do**

    $j \leftarrow j+1$

    $\hat{\Delta}(j) \leftarrow$ minimize (1.95) w.t.r to $(t_{end} - t_{start})/2$ ▷ Evaluate the optimal Δ at every time

    $Y_{j+1} \leftarrow Y_j + \mu(t_{start})\hat{\Delta}(j) + \sigma(t_{start})\sqrt{\hat{\Delta}(j)}\mathcal{N}(0,1)$ ▷ Set Y_{j+1}

    $t_{start} \leftarrow t_{start} + \hat{\Delta}(j)$ ▷ increment t_{start};

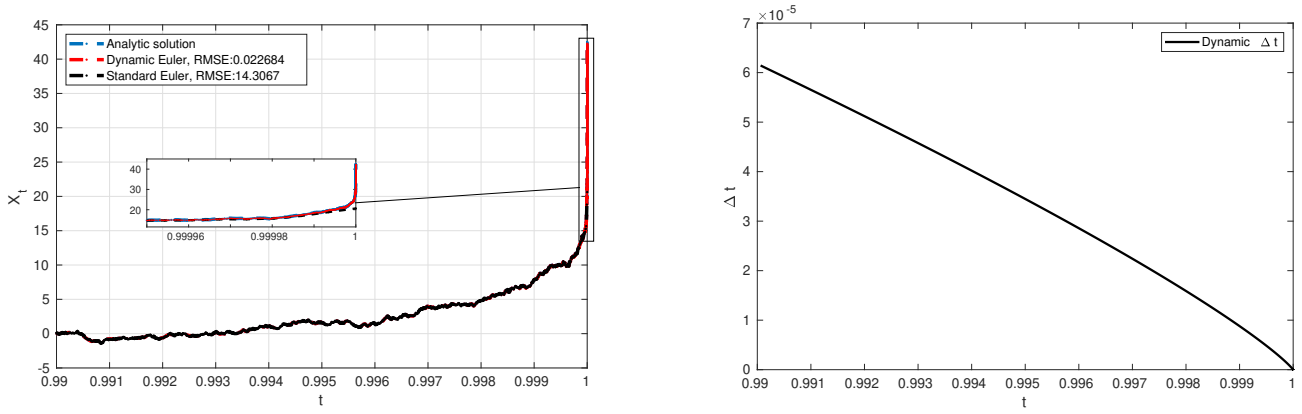
  return $Y_{j+1}, \hat{\Delta}(j)$

From the algorithm (1), we notice that the discretization step Δ is chosen dynamically at every iteration during the simulating of a toy model.

To test how accurate our method can simulate the trajectory of a toy model, we compare the toy model trajectory obtained with the dynamic Euler method in Eq. (1.104) to those obtained with the analytic solution in Eq.(1.85) and the Euler Maruyama method in Eq.(1.86). To implement the algorithm 1, we set minRMSE, $\alpha, \beta, c_\sigma, c_\mu, t_{start}$ and t_{end}. Here, we set minRMSE $= 5 \times 10^{-8}, \alpha = 0.999, \beta = 0.25, c_\sigma = 10, c_\mu = 2, t_{start} = 0.99$ and $t_{end} = 1 - 6 \times 10^{-11}$. To simulate the approximate solution of a toy model with a standard Euler-Maruyama method, the discretization step is set to be equal to $\Delta t_i = (t_{end} - t_{start})/\max_j$, where $\max_j$ is a maximum number of iteration of the dynamic Euler method.

In Figures 1.3(a), we plot the trajectories of a toy model simulate using Dynamic Euler method (red-line), analytic formula (blue-line) and Euler Maruyama method (black-line). Figure 1.3(b) shows the trajectory of the dynamic distrization step. Note that the processes stop at iteration 1532 and therefore $\max_j = 1532$. Not-surprisingly, the trajectories of a toy model simulated with the Euler Maruyama method and the analytic solution are significantly distinct around the exploding time (see Figure 1.3(a)). This distinction is further confirmed by looking at the value of RMSE as the latter is almost 14%, which is very large. On the other hand, the trajectory of a toy model simulated with the dynamic Euler method is virtually identical to the one generated with the analytic formula as we can see from Figure 1.3(a). We also observe that the RMSE is very small, 0.022%. Figure 1.3(b) shows that the estimated discretization step $\hat{\Delta} t$ becomes very small as the time t gets close to the exploding time t_{end}. These results indicate that the dynamic Euler method helps us to simulate more accurately the trajectories of a toy model.

Figure 1.3 In left-hand Figure 1.3(a), we compare the explicit solution (red line) and its approximate ones obtained with the dynamic Euler algorithm (blue line) and the standard Euler algorithm (black line). The sub-figure plots the comparison around the exploding point. In right-hand Figure 1.3(b), we plot the dynamic discretization step $\hat{\Delta}$. The results show that dynamic Euler discretization accurately approximates the explicit solution even around the exploding point while standard Euler algorithm fails.



(a) (b)

1.5 Application of dynamic Euler method to exploding models

In this section, we use the dynamic Euler method introduced above to simulate the trajectories of the exploding models. Note that the exploding models are different from the toy of an exploding model in Eq. (1.84). To implement the dynamic Euler method, we need to estimate the discretization step at each step of the iteration to obtain the desired or target RMSE. That is equivalent to solve problem (1.95). On the other hand, to solve problem (1.95), we require to compute the expected value $\mathbb{E}[(\Delta Y_i^{Eul} - \Delta Y_i)^2]$. Contrary to the case of a toy model which the analytical solution of a toy model, the analytical solutions of the exploding models are not available. Therefore, we cannot explicitly compute the expected value ($\mathbb{E}[(\Delta Y_i^{Eul} - \Delta Y_i)^2]$), and the problem (1.95) cannot be solved. However, we can rely on a toy model to circumvent the problem. We proceed as follows:

1. We consider the expected value in Eq. (1.93) since the latter depends only on the time process.

2. We express the time process (t) of a toy model in Eq. (1.84) as the function of the stochastic process Y of exploding models in Eq. (1.82).

3. We replace the time process (t) in Eq. (1.84) with the new one which depends on the stochastic process of exploding models, e.i. $t(Y)$.

4. To simulate the trajectories of exploding models in Eq. (1.82) using the dynamic Euler method, we then use the new expected value in Eq. (1.84) instead of computing the expected value ($\mathbb{E}[(\Delta Y_i^{Eul} - \Delta Y_i)^2]$) to solve problem in Eq. (1.95).

Let us now discuss how to express the time process of a toy model in Eq. (1.84) in the function of the stochastic process (Y) of exploding models in Eq. (1.82). To do so, we assume the drift and diffusion coefficients of exploding process in Eq. (1.82) explode at the same magnitude as the ones of a toy model in Eq. (1.84). Equivalently, we set the absolute values of the drift and diffusion coefficients of exploding models in Eq. (1.82) are equal that of a toy model in Eq.(1.84):

$$c_\mu \left(\frac{1}{(t_{end} - t_\mu)^\alpha} - 1 \right) = |\mu(Y_t)| \tag{1.97}$$

$$c_\sigma \left(\frac{1}{(t_{end} - t_\sigma)^\beta} - 1 \right) = |\sigma(Y_t)|. \tag{1.98}$$

Solving (1.97) and (1.98), and we get:

$$t_\mu(Y_t) = t_{end} - \left(\frac{c_\mu}{c_\mu + |\mu(Y_t)|} \right)^{\frac{1}{\alpha}} \tag{1.99}$$

$$t_\sigma(Y_t) = t_{end} - \left(\frac{c_\sigma}{c_\sigma + |\sigma(Y_t)|} \right)^{\frac{1}{\beta}}. \tag{1.100}$$

We can see that the times t_μ and t_σ depend on the stochastic process $Y_t$of the exploding model in Eq. (1.82). Substituting the obtained times in Eq. (1.99) and (1.100) in Eq. (1.93), and we now get:

$$
\begin{aligned}
D(\Delta) &= c_\mu^2 \left(\frac{1}{(t_{end} - t_\mu(Y_i))^\alpha} \Delta t_i - \frac{1}{1-\alpha} (t_{end} - t_\mu(Y_{i+1}))^{-\alpha+1} + \frac{1}{1-\alpha} (t_{end} - t_\mu(Y_i))^{-\alpha+1} \right)^2 \\
&+ c_\sigma^2 \left[\frac{1}{(t_{end} - t_\sigma(Y_i))^\beta} - 1 \right]^2 \Delta t_i - c_\sigma^2 \frac{1}{1-2\beta} (t_{end} - t_\sigma(Y_{i+1}))^{-2\beta+1} + c_\sigma^2 \Delta t_i \\
&+ c_\sigma^2 \frac{1}{1-2\beta} (t_{end} - t_\sigma(Y_i))^{-2\beta+1} + c_\sigma^2 \frac{2}{1-\beta} (t_{end} - t_\sigma(Y_{i+1}))^{-\beta+1} - c_\sigma^2 \frac{2}{1-\beta} (t_{end} - t_\sigma(Y_i))^{-\beta+1} \\
&- 2c_\sigma^2 \left[\frac{1}{(t_{end} - t_\sigma(Y_i))^\beta} - 1 \right] \left[- \frac{1}{1-\beta} (t_{end} - t_\sigma(Y_{i+1}))^{-\beta+1} + \frac{1}{1-\beta} (t_{end} - t_\sigma(Y_i))^{-\beta+1} - \Delta t_i \right] - \text{RMSE}^*
\end{aligned}
\tag{1.101}
$$

$$
\begin{aligned}
&= c_\mu^2 \left(\frac{c_\mu + |\mu(Y_i)|}{c_\mu} \Delta t_i - \frac{1}{1-\alpha} \left(\frac{c_\mu}{c_\mu + |\mu(Y_{i+1})|} \right)^{\frac{-\alpha+1}{\alpha}} + \frac{1}{1-\alpha} \left(\frac{c_\mu}{c_\mu + |\mu(Y_i)|} \right)^{\frac{-\alpha+1}{\alpha}} \right)^2 \\
&+ (|\sigma(Y_i)|^2 + c_\sigma^2) \Delta t_i - \frac{c_\sigma^2}{1-2\beta} \left(\frac{c_\sigma}{c_\sigma + |\sigma(Y_{i+1})|} \right)^{\frac{-2\beta+1}{\beta}} + \frac{c_\sigma^2}{1-2\beta} \left(\frac{c_\sigma}{c_\sigma + |\sigma(Y_i)|} \right)^{\frac{-2\beta+1}{\beta}} \\
&+ \frac{2c_\sigma^2}{1-\beta} \left(\frac{c_\sigma}{c_\sigma + |\sigma(Y_{i+1})|} \right)^{\frac{-\beta+1}{\beta}} - \frac{2c_\sigma^2}{1-\beta} \left(\frac{c_\sigma}{c_\sigma + |\sigma(Y_i)|} \right)^{\frac{-\beta+1}{\beta}} \\
&- 2c_\sigma |\sigma(Y_i)| \left(- \frac{1}{1-\beta} \left(\frac{c_\sigma}{c_\sigma + |\sigma(Y_{i+1})|} \right)^{\frac{-\beta+1}{\beta}} + \frac{1}{1-\beta} \left(\frac{c_\sigma}{c_\sigma + |\sigma(Y_i)|} \right)^{\frac{-\beta+1}{\beta}} - \Delta t_i \right) - \text{RMSE}^*
\end{aligned}
\tag{1.102}
$$

The expected value in Eq. (1.102) depends now on the stochastic process Y of exploding models in Eq.(1.82). Suppose that the drift and diffusion coefficients of exploding models in Eq. (1.82) given, to determine the dynamic discretization step $\hat{\Delta}t$ at iteration j, we solve the following problem:

$$
\hat{\Delta}t_j = \underset{\Delta t_j}{\arg\min} D(\Delta). \tag{1.103}
$$

where $\Delta_j = (t_{end} - tj)/2$. Using the estimated discretization step $\hat{\Delta}t_j$, the approximate solution Y of exploding models in Eq. (1.82) by dynamic Euler method, denote by $\widetilde{Y} = \{\widetilde{Y}_t, t \in [0,T]\}$ and defines as:

$$
\widetilde{Y}_{j+1} = \widetilde{Y}_j + \mu(\widetilde{Y}_j)\hat{\Delta}t_j + \sigma(\widetilde{Y}_j)\Delta W_j, \quad j = 0,1,2,\ldots, \quad \text{and} \quad \widetilde{Y}_0 = Y_0. \tag{1.104}
$$

On the other hand, as the analytical solution of exploding models in Eq. (1.82) is not available, we propose the "quasi-exact" solution of exploding models (1.82). The quasi-exact solution of Y is obtained

by sub-discretizing the approximated solution of Y obtained with dynamic Euler discretization $\widetilde{Y}$. Suppose that $\bar{t}_i = \frac{i\hat{\Delta}t_j}{N}; \quad i = 0,\ldots,N$ is a sub-discretization of the time interval $[0,\hat{\Delta}(i)]$ with discretization step $\hat{\Delta}(i)$ is estimated from the equation (1.95) and N the number of substeps of the quasi exact solution. We set $\Delta W_i = W_{\bar{t}_{i+1}} - W_{\bar{t}_i}, \quad i = 0,\ldots,N-1$ where ΔW_i is the increment of the Wiener process with independent and identically distributed normal random variables with mean zero and variance $\frac{\hat{\Delta}t_j}{N}$. Let us denote by $\hat{Y} = \{\hat{Y}_{\bar{t}}, \bar{t} \in [0,\hat{\Delta}]\}$ the quasi-exact solution and satisfying the following recursive equation

$$\hat{Y}_{i+1} = \hat{Y}_i + \mu(\hat{Y}_i)\frac{\hat{\Delta}t_j}{N} + \sigma(\hat{Y}_i)\Delta W_i, \tag{1.105}$$

for $i = 0,\ldots,N-1$ with initial value $\hat{Y}_0 = Y_0$. Below, we present the algorithm to simulate the trajectories of exploding model in Eq. (1.82) using the dynamic Euler discretization in Eq (1.104) and the quasi-exact solution in Eq. (1.105).

Algorithm 2 : Algorithm to simulate the trajectories of exploding models based on dynamic Euler method

procedure ITERATION($N, Y_0, \mu(\cdot), \sigma(\cdot)$) ▷ Initialize all input parameters
  $Z_0 \leftarrow$ Starting point or Initial price
  $z_0 \leftarrow$ Exploding point
  $N \leftarrow$ Number of substeps for almost true Euler Solution
  $\hat{Y}_0 \leftarrow Y_0$ ▷ Initialize the almost true Euler $\mu(\cdot)$
  $t(0) \leftarrow 0$ ▷ Initialize the exploding time
  $j \leftarrow 0$
  while $(\widetilde{Z}_j > z_0)$ **do**
    $j \leftarrow j+1$
    $t_\mu(Y_j)$ and $t_\sigma(Y_j)$ ▷ Calculate the time using the Eq. (1.99) and (1.100)
    $t_\mu(j) = \min(t_{end}, t_\mu(j))$ and $t_\sigma(j) = \min(t_{end}, t_\sigma(j))$; ▷ Keep the time t_{end}
    $t_{\max}(j) = \max(t_\mu(j), t_\sigma(j))$; ▷ Take the maximum between t_μ and t_σ
    $\hat{\Delta}(j) \leftarrow$ (1.95) w.t.r to $(t_{end} - t_{\max}(j))/2$ ▷ Evaluate the optimal Δ at each iteration
    $\widetilde{Z}_{j+1} \leftarrow \widetilde{Z}_j + \mu(\widetilde{Z}_j)\hat{\Delta}(j) + \sigma(\widetilde{Z}_j)\sqrt{\hat{\Delta}(j)}\mathcal{N}(0,1)$ ▷ Increment of dynamic Euler $\widetilde{Y}_{j+1}$
    $t(j+1) \leftarrow t(j) + \hat{\Delta}(j)$ ▷ Increment of the exploding time t_{j+1};
    ▷ Approximate the almost exact solution of the exploding model.
    $X \leftarrow \hat{Z}_j$
    for $k \leftarrow 1$ to N **do**
      $X \leftarrow X + \mu(X)\frac{\hat{\Delta}(j)}{N} + \sigma(X)\sqrt{\frac{\hat{\Delta}(j)}{N}}\mathcal{N}(0,1)$
    $\hat{Z}_{j+1} \leftarrow X$ ▷ Increment of quasi-exact solution $\hat{Z}_{j+1}$
  return $\widetilde{Z}_{j+1}, \hat{Z}_{j+1}, t_{j+1}, \hat{\Delta}(j)$

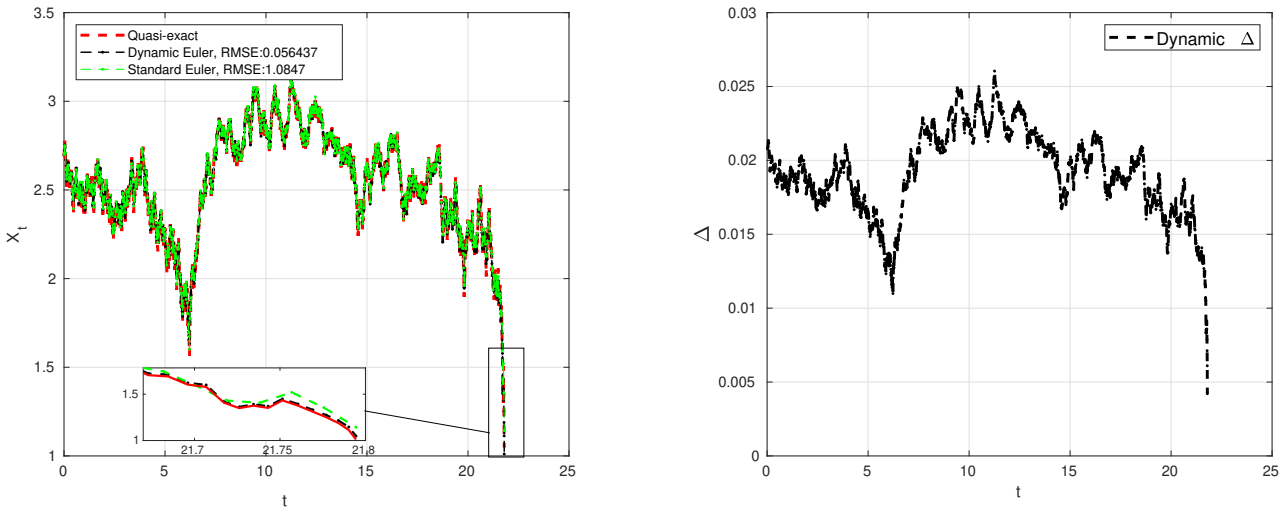
To asses how accurate the proposed algorithm 2 simulates the trajectories of exploding model, we consider the exploding model introduced in Section 1.2.4:

$$dY_t = \frac{a\sigma^*[k - \ln(\bar{Y} - y_0)]^{\frac{a}{\sigma^*}}}{2(Y_t - y_0)(k - \ln(Y_t - y_0))^{\frac{a}{\sigma^*}+1}}dt + \frac{\sigma^*(k - \ln(\bar{Y} - y_0))^{\frac{a}{\sigma^*}}}{(k - \ln(Y_t - y_0))^{\frac{a}{\sigma^*}}}dW_t, \quad \bar{Y} = Y_0 > y_0 > 0 \tag{1.106}$$

where y_0 is an exploding point, and $\bar{Y}$ is the point before the explosion. The parameters $\sigma^* = 0.271335, a = -0.295774, k = 1.646359$ are calibrated from the flash crash data of SPY of a 06/05/2010 (see Section 1.3). We simulate the trajectories of the exploding model in Eq. (1.106) using the dynamic Euler discretization in Eq. (1.104), the Euler-Maruyama method and the quasi-exact solution in Eq. (1.105). To implement the algorithm 2, we set minRMSE $= 1 \times 10^{-7}, \alpha = 0.99, \beta = 0.3, c_\sigma = 0.01, c_\mu = 0.01, N = 1000; \bar{Y} = 2.7; y_0 = 1, t_{end} = 1 - 10^{-8}$ and $Y_0 = 2.7$. To use Euler-Maruyama method, we set the discretization step equal to $\Delta = (t(end) - t(0))/\max_j$, where $\max_j$ is a maximum number of iteration and the dynamic Euler method and t_{end} the exit time of the dynamic Euler method. We simulate 1000 replicate trajectories of the exploding model (1.106) using the three methods.

In Figure 1.4(a), we plot one trajectory of exploding model in Eq. (1.106) using dynamic Euler method in Eq. (1.104) (black-line), quasi-exact solution in Eq. (1.105) (red-line) and Euler Maruyama method (green-line). In Figure (1.3(b)), we plot the trajectory of the estimated distrization step. Figure 1.4(a) shows the trajectory of exploding model in Eq. (1.106) simulated with Euler Maruyama method is significantly distinct to the true one generated with the quasi-exact solution. On the other hand, the trajectory of the exploding model in Eq. (1.106) simulated with dynamic Euler method is virtually identical to that of the quasi-exact solution (as also shown in sub-figure 1.4(a)). Furthermore, the RMSE between the dynamic Euler method and the quasi exact solution is 0.05%, which is very small compared to the one obtained between the Euler Maruyama method and the quasi-exact solution, 1.08%. Figure. 1.3(b) shows that the discretization step becomes very small when approaching the exploding point. We also observe the trajectory of the estimated discretization step has a similar behaviour as the trajectory of exploding model. That means dynamic Euler method allows us to choose the discretization step according to the shape of the trajectory.

Figure 1.4 Left-hand Figure 1.3(a): plots the trajectories of exploding model in Eq. (1.106) simulated with the quasi-exact solution (red-line), dynamic Euler method (black line) and the Euler Maruyama method (green line). Sub-figure compares the trajectories in the exploding region. Right-hand Figure 1.3(b): plots the trajectory of the estimated discretization step $\hat{\Delta}$.



(a) (b)

Table (1.5) reports the results of the RMSE and the exploding times. We only report the results of 10 replications amongst 1000. From the first column in Table (1.5), we see that the RMSE of the dynamic Euler method is small than to the RMSE of the Euler Maruyama method for all replications. These simulations confirm further the dynamic Euler method simulated the trajectories of the exploding model more accurately than the Euler Maruyama method. In the second column of Table (1.5), we observe that the trajectories of the exploding model simulated with the dynamic Euler method explode in the same that as the trajectories of exploding model simulated with the Euler Maruyama method for all replications. That means the trajectories of exploding model simulated with both dynamic Euler method and Euler Maruyama method explode at the same points.

Table 1.5 The results of the RMSE obtain by comparing the the exact solution of the exploding process and its approximate solution obtain with the dynamic Euler algorithm and the standard Euler algorithm for 10 replicate trajectories. The results show that the RMSE obtain with dynamic Euler algorithm are very small compared to the ones obtain with the standard Euler algorithm for all 10 replications. We shall also note that both solutions explode exactly at the same time.

Iteration	RMSE(%)		Exploding times	
	Dynamic Euler	Standard Euler	Dynamic Euler	Standard Euler
1	0.1498	1.2229	42.3372	42.3372
2	0.2905	1.9789	2.7462	2.7462
3	0.1310	1.7560	12.8066	12.8066
4	0.2049	1.5313	4.3286	4.3286
5	0.0490	0.8452	197.5121	197.5121
6	0.0706	0.6685	347.0933	347.0933
7	0.0729	1.3352	18.6001	18.6001
8	0.0298	0.5610	718.999	718.999
9	0.1941	1.8051	3.7221	3.7221
10	0.2060	1.5673	10.57601	10.57601

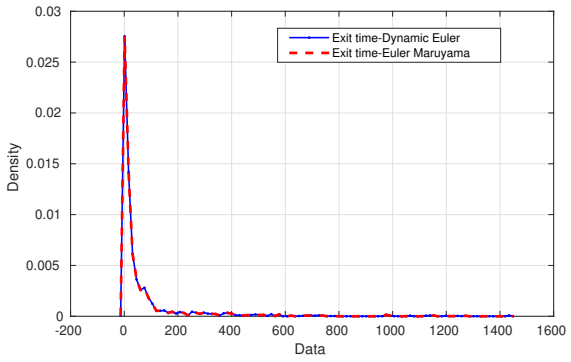
In Figures 1.5(a) and 1.5(b), we plot the distributions of the RMSE and the exit time for all 1000 replications. Figure 1.5(a) shows the distributions of the exit time of the dynamic Euler (blue-line) and the distribution of the Euler Maruyama method (red-line). We observe the dynamic Euler method and the Euler Maruyama method explode exactly at the same time since the two distributions are similar. This result suggests that the trajectories of the exploding model simulate with the dynamic Euler method explode exactly at the same point as the ones simulate with the Euler Maruyama method. Figure 1.5(b) presents the density distribution of the $\text{RMSE}_{\text{Dynamic Euler}}$[4] and the density distribution of the $\text{RMSE}_{\text{Euler Maruyama}}$.[5] From Figure 1.5(b), we can see that the two density distributions are different. The distribution of the $\text{RMSE}_{\text{Dynamic Euler}}$ has a peak that is too high compared to the density distribution of the $\text{RMSE}_{\text{Euler Maruyama}}$. Furthermore, we can see the upper tail of the density distribution of the $\text{RMSE}_{\text{Dynamic Euler}}$ is very short, while that of the density distribution of the

[4] $\text{RMSE}_{\text{Dynamic Euler}}$ is the RMSE between the dynamic Euler method and quasi-exact solution

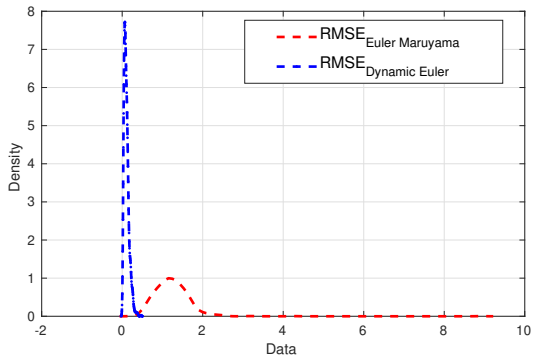
[5] $\text{RMSE}_{\text{Euler Maruyama}}$ is the RMSE between the the Euler-Maruyama method and quasi-exact solution

$\text{RMSE}_{\text{Euler Maruyama}}$ is very larger. These results further confirm the RMSE of the dynamic Euler method is always small than the one of Euler Maruyama method.

Figure 1.5 Left-panel: Figure 1.5(a) plots the density distribution of exit times of the Euler Maruyama method (red-line) and that of the dynamic Euler method (blue-line) for 1000 replicated trajectories. Right-panel: Figure 1.5(b) plots the density distributions of the RMSE of the dynamic Euler method (blue line) and the distribution of the RMSE of the Euler Maruyama method (red line).



(a)



(b)

We conclude that one can rely on the dynamic Euler method to simulate more accurately the trajectories of exploding models.

1.6 Conclusion

In this chapter, we proposed a model generating drift bursts. We showed that the feedback effect is one potential reason for short-lived explosive trends in the path of financial market prices. In Section 1.2, we discussed the existence and uniqueness of the solution of the SDE with exploding coefficients. We showed in Theorem 12 that the SDE with exploding coefficients admits a weak solution, which is unique in the sense of probability distribution and defined up to the explosion time. We also discussed some of the examples of the models generating drift burst (exploding models). We showed the drift and diffusion coefficients for each discussed example of exploding models, explode at a single finite point of $\mathbb{R}$, with the drift coefficient explode at the magnitude higher that of the diffusion coefficient. In Section 1.3, we estimated the parameters of exploding models by fitting each model to the E-mini S&P 500 flash crash of May 2010(SPY). In Section 1.4, we proposed

the dynamic Euler method, which allows us to simulate more accurately the trajectories of exploding models. A Monte Carlo simulations showed confirmed that the dynamic Euler method simulated the trajectories of the exploding model more accurately than the Euler Maruyama method.

We should note that some work can be done here. For example, we can extend this work to derive the probability distribution of the exploding time in similar ways as in Karatzas and Ruf (2016).

Chapter 2

Nonparametric estimation of exploding diffusion function

2.1 Nonparametric estimation: literature review

The estimation of the diffusion coefficient is an important subject of interest in financial economics as the latter plays a significant role in finance applications. For example, the diffusion coefficient can be used, to model option prices, term structure of interest rate, so forth (see Baxter and Ronnie 1996 for an introduction of some these applications). The non-parametric identification of diffusion coefficient from continuous data has been considered in the literature, see, for example, Prakasa Rao (1999). However, many authors have pointed out that the assumption that sampling observations are continuous is too restrictive. Because in reality, it is impossible to observe a process continuously over a given interval, maybe because some data may not be available at every point of time (Nicolau 2003a and Prakasa Rao 1999). Therefore, it is reasonable that in recent years many authors focused on the estimation of the diffusion process based on discrete-time observation. Florens-Zmirou (1993) proposed a non-parametric estimator of the diffusion process based on discrete-time observations. Jiang (1998) extended the work of Florens-Zmirou (1993) using a Gaussian kernel and then proposed a drift estimator based on a diffusion coefficient estimator and the stationary density. Stanton (1997) also used the Florens-Zmirou (1993) estimators to estimate short-term interest rates model. Ait-Sahalia (1995) proposed a semiparametric estimator of a diffusion coefficient based on a parametric drift coefficient, while

Bandi and Phillips (2003) proposed a more general estimator of the diffusion and drift coefficients under the wide the assumption on DGP. The non-parametric estimator of Bandi and Phillips (2003) is similar to that one of Florens-Zmirou (1993). However, in Bandi and Phillips (2003), the existence of marginal density is not required and therefore, the stationary condition is indeed not needed. Renò (2008) proposed a non-parametric estimator of a diffusion coefficient based on Fourier analysis of the state variable trajectories observed and on the estimation of quadratic variation between observation by mean of the realized volatility. The quoted literature deals with the case of the bounded coefficients. Our objective is the estimation of explosive behaviour like that observed, for example, during the flash crash of May 2010. That means, differently from the quoted literature, and in line with previous Chapter 1, we will allow the drift and volatility coefficients to explode, and provide a non-parametric estimator of the diffusion coefficient in this particular case.

The proposed nonparametric estimator is constructed based on the theory of Nadaraya-Watson estimator of the form

$$S^n(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h_{n,\bar{T}}}\right)\bar{\sigma}^2(t_i)}{\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h_{n,\bar{T}}}\right)}, \tag{2.1}$$

where $\bar{\sigma}^2$ is a consistent estimator of the spot volatility at time t_i while $h_{n,\bar{T}}$ is a bandwidth parameter (Härdle 1990). The implementation of a non-parametric estimator is a two-steps procedures. First, we estimate the time series $\hat{\sigma}(t_i) = \frac{(X_{t_{i+1}}-X_{t_i}-\hat{\mu}(X_{t_i})\Delta)^2}{\Delta}$ where $\hat{\mu}(\cdot)$ is a suitable drift estimator and $\Delta = \frac{\bar{T}}{n}$.[1] We then implement$\hat{\sigma}(t_i)$ in the second step using equation (2.1). Since a drift coefficient explodes at a magnitude higher than the one of a diffusion coefficient during the flash crash. Thus, including a non-parametric estimator of a drift coefficient in spot volatility estimator allow us to improve the accuracy on estimating the exploding diffusion process.

The main contribution of this work is the following. We introduce the new nonparametric estimator of the form (2.1) for the stochastic differential equation and, we then derive the asymptotic properties of the proposed estimator. It is well- known, in principle, to provide the consistency of nonparametric estimator of diffusion processes we do not need to impose a condition $t_i \to \infty$ (see Theorem 6 Bandi and Phillips 2003). However, we only need the existence of solution of stochastic differential equation holds, and the regularities conditions

[1] We refer the reader to the Section (2.2) for more details on how the nonparametric estimator is constructed

on the drift and diffusion coefficients also hold. We borrow from the limit theory of the convergence of a semimartingale to a process with independent increment to asses the asymptotic properties of the estimator. We consider the popular Florens-Zmirou estimator as the benchmark. Then, we compare a newly proposed nonparametric estimator to Florens-Zmirou estimator by mean of Monte Carlo simulations. To asses the unbiased nature of the proposed estimator, we simulate the discrete data using two types of models in a small sample. We simulate the data with the CEV model and the exploding models. With the exploding models, we can generate flash crash prices. The Monte Carlo results based on the data generated with the CEV model show that the results obtained with the proposed estimator and Florens-Zmirou (FZ) estimators are almost similar. Thus, the correction does not contribute to this case. On the other hand, the Monte Carlo results based on the flash crash data generate with the exploding model, show that the proposed estimator estimates more accurately the exploding diffusion coefficient compared to the FZ estimator, especially in exploding region. Finally, we estimate the diffusion coefficient of the exploding model from the E-mini S&P 500 flash crash data of May 2010. The result shows that in the interval $4.7-4.6738$ where the drift coefficient $|\mu(x)|$ is higher, the proposed estimator shows a significant improvement over the FZ estimator, especially around the explosive price.

2.1.1 Model and Setting

The stochastic differential equation under consideration is,

$$\begin{cases} dX_t & = \mu(X_t)dt + \sigma(X_t)dW_t \quad X_0 = \bar{X} \in \mathbb{R}, \\ \sigma(x) & = \sigma^* e^{\frac{2}{\sigma^*}\int_{\bar{X}}^{X} g(u)du} \\ \mu(x) & = \sigma^* g(x) e^{\frac{2}{\sigma^*}\int_{\bar{X}}^{X} g(u)du} \end{cases} \tag{2.2}$$

defined over the time interval $[0,T]$, in the filtered probability space $(\Omega, \mathscr{P}, (\mathscr{F}_t)_{t\geq 0}, \mathscr{F})$ satisfying the usual conditions and $\{W_t\}_{t\in[0,T]}$ is a standard Brownian. The terms σ^* and $\bar{X}$ are both constant, with $\bar{X}$ is considered to be a price before a flash crash starts. The initial conditionX_0 is square-integrable and is taken to be independent of $\{W_t\}_t \in [0,T]$ and measurable with respect to $\mathscr{F}_0$. The unknown function $g(\cdot)$ satisfies the conditions of the

assumption (3) in Chapter 1. Here, we present some example of SDE in Eq. (2.2) discussed in Chapter 1.

$$dX_t = \frac{a_1\sigma_1^*[k-\ln(\bar{X}-x_0)]^{\frac{a_1}{\sigma_1^*}}}{2(X_t-x_0)(k-\ln(X_t-x_0))^{\frac{a_1}{\sigma_1^*}+1}}dt + \frac{\sigma_1^*(k-\ln(\bar{X}-x_0))^{\frac{a_1}{\sigma_1^*}}}{(k-\ln(X_t-x_0))^{\frac{a_1}{\sigma_1^*}}}dW_t, \tag{2.3}$$

$$dX_t = -\frac{a\sigma^*(k-a\ln(X_t-x_0))^{\frac{1}{\sigma^*}-1}}{2(X_t-x_0)(k-a\log(\bar{X}-x_0)^{\frac{1}{\sigma^*}}}dt + \frac{\sigma^*(k-a\ln(X_t-x_0))^{\frac{1}{\sigma^*}}}{(k-a\log(\bar{X}-x_0)^{\frac{1}{\sigma^*}}}dW_t, \quad X_0=\bar{X}>x_0, t\in[0,\bar{T}], \tag{2.4}$$

where $a,\sigma^*,k,a_1,\sigma_1^*$ need to be estimated.The initial value $X_0=\bar{X}>x_0$. The term x_0 is a point in which the trajectory of the solution process of the SDE in Eq. (2.2) explodes. As we can, when $X\mapsto x_0$, the drift and diffusion coefficients in Eq. (1.71) and (2.4) diverge to infinity, but the drift coefficient explodes with a magnitude higher than that of the diffusion coefficient.

We aim to estimate the exploding diffusion coefficient $\sigma(\cdot)$ of the SDE in Eq. (2.2) given the observe discrete data of $X_t=\{X_0,X_1,\ldots,X_n\}$ with $t_0=0<t_1<\cdots<t_n=T$ where $t_i=\frac{iT}{n}$ on the time interval $[0,T]$. To do so, we need to propose a nonparametric estimator of this kind of diffusion coefficient. To derive a consistent diffusion estimator, we need to verify if the drift and diffusion coefficients satisfy the regularities conditions in $\mathbb{R}$, namely the (local) Lipschitz condition and linear growth conditions (Florens-Zmirou 1993, Bandi and Phillips 2003, Renò 2008). However, the above-stated conditions do not hold in our case, as shown in Proposition 1 of Chapter 1, that both drift and diffusion coefficients explode at a single point of $\mathbb{R}$, i.e. $\sigma(x)$ and $\mu(x)$ diverge to infinity when x approachesx_0. On the other hand, we know that the process X_t stops once it hits the exploding point. However, we are only interested in working in the domain where the process X_t is still ongoing, namely, on the subset of $\mathbb{R}$ which does not contain an exploding point.

Suppose that x_0 is a finite unique singular point of $\mathfrak{I}=(1,r)\subseteq\mathbb{R}$, and the function $g(\cdot)$ diverges to infinity as x approaches x_0, e.i. $g(x_0)=\pm\infty$. For instance, assuming that $g(x)=\frac{1}{(x-x_0)^2}$ and we can see that the function g explodes at x_0. Now, assuming that $\mathfrak{D}$ is a subset of $\mathbb{R}\setminus\{0\}\subseteq\mathbb{R}$ and, we can prove that g is twice continuously differentiable on $\mathfrak{D}$. Consequently, we can show that the drift $(\mu(\cdot))$ and diffusion $\sigma(\cdot)$ coefficients satisfy the local Lipschitz and linear growth conditions for every x on $\mathfrak{D}$ (see the Proposition (1) in Chapter 1). To build a

consistent estimator of the exploding diffusion coefficient, we must work on a subset $\mathfrak{D}$. We define $\mathfrak{D}$ as:

$$\mathfrak{D} = [x_1, r) \subset (x_0, \mathrm{r}) \subset \mathfrak{I}, \tag{2.5}$$

where $-\infty \leq l < x_0 < x_1 < \mathrm{r} \leq \infty$. Let assume that $S = \inf\{t \geq 0 : X_t \notin \mathfrak{I}\}$ is the exit time of the process (X) once it hits an exploding point x_0 of $\mathfrak{I}$. The time before the trajectory (X) of SDE in Eq. (2.2) reaches the exploding point x_0, denote by $\bar{T}$, it sets as $\bar{T} < S \leq T$. Hence, the process X does not explode over the time interval of $[0, \bar{T}]$.

In the remainder of Chapter, we now consider the following SDE:

$$\begin{cases} dX_t & = \mu(X_t)dt + \sigma(X_t)dW_t \quad X_0 = x \in \mathfrak{D}, \\ \sigma(x) & = \sigma^* e^{\frac{2}{\sigma^*}\int_X^x g(u)du} \\ \mu(x) & = \sigma^* g(x) e^{\frac{2}{\sigma^*}\int_X^x g(u)du} \end{cases} \tag{2.6}$$

defined over the time interval $[0, \bar{T}]$, in the filtered probability space $(\Omega, \mathscr{P}, (\mathscr{F}_t)_{(0 \leq t \leq \bar{T})}, \mathscr{F})$ satisfying the usual conditions and $\{W_t\}_{t \in [0,\bar{T}]}$ is a standard Brownian. The initial condition x is still square integrable and is taken to be independent of $\{W_t\}_t \in [0, \bar{T}]$ and measurable with respect to $\mathscr{F}_0$.

Remark 7. *Note that the SDE in Eq (2.6) and (2.2) are different. In the SDE (2.2), the solution X explodes over the time interval of $[0,T]$, and the drift and diffusion coefficients do not satisfy the local Lipschits and linear growth conditions for every x in $\mathbb{R}$. However, the solution X of SDE in Eq. (2.6) is continuous in the time interval of $[0, \bar{T}]$. Furthermore, the drift and diffusion coefficients in Eq- 1 satisfy local Lipschits and linear growth conditions for every x in $\mathfrak{D}$. It is therefore reasonable that we consider the SDE in Eq. (2.6) for the remainder of the work, since its coefficients satisfy the local Lipschitz condition and linear growth conditions.*

Assumption 3.

Consider the SDE (2.6). Assume that following conditions hold.

(i) $g \in C^2(\mathfrak{D})$.

(ii) The functions $\sigma(\cdot)$ and $\mu(\cdot)$ are time-homogeneous and $\mathfrak{B}-$ measurable functions on $\mathfrak{D}$ where $\mathfrak{B}$ be a the σ-field generated by Borel sets on $\mathfrak{D}$. Using condition (i), then we assume the functions $\mu(\cdot)$ and $\sigma(\cdot)$ are

atleast twice continuously differentiable, and hence they satisfy (local) Lipschitz and growth conditions. Thus, for every compact subsets $\mathfrak{J}$ of $\mathfrak{D}$ there exits two positive constants K_1 and K_2 such that, for all $x, y \in \mathfrak{J}$,

$$|\sigma(x)-\sigma(y)|+|\mu(x)-\mu(y)| \leq K_1|x-y|; \tag{2.7}$$

and

$$|\sigma(x)|+|\mu(x)| \leq K_2(1+|x|^2). \tag{2.8}$$

(iii) Suppose there exists a constant $M>0$ such that $|\sigma(x)| \leq M$ and $|\mu(x)| \leq M$ for all x;

Assumption 3.(ii) implies the existence and the uniqueness of a strong solution for SDE (2.6) over time interval $[0, \bar{T}]$ holds. We also note that the assumption 3.(i)–(iii) allow us to write the infinitesimal moments of the changes in the solution to (2.6) in the terms of $\mu(\cdot)$ and $\sigma(\cdot)$. Further,

$$M^1(x) = \lim_{\Delta \longrightarrow 0} \frac{1}{\Delta} E[X_{t+\Delta} - X_t | X_t = x] = \mu(x) \tag{2.9}$$

$$M^2(x) = \sigma^2(x) = \lim_{\Delta \longrightarrow 0} \frac{1}{\Delta} E[(X_{t+\Delta} - X_t)^2 | X_t = x] = (\sigma^*)^2 e^{\frac{4}{\sigma^*}\int_X^x g(u)du}. \tag{2.10}$$

Assumption 4. *The kernel function $K(\cdot)$ is assumed to be continuously differentiable function (it is bounded and Lipschitz-continuous), symmetric and non decreasing with a bounded derivative $K'(\cdot)$. It also satisfies the following conditions*

$$\int_{\mathbb{R}} K(s)ds = 1, \quad \int_{\mathbb{R}} sK(s)ds = 0, \quad \int_{\mathbb{R}} K^2(s)ds < \infty, \quad \sup K(s) < C_3$$

and

$$K_t = \int_{\mathbb{R}} s^2 K(s)ds < \infty.$$

We also provide the following assumption for a sequence of bandwidths $h_{n,\bar{T}}$ as follows:

Assumption 5.

(i) For all $n \in \mathbb{N}, h_{n,\bar{T}}$ is a real sequence such that, $h_{n,\bar{T}} \to 0$, $nh_{n,\bar{T}} \to \infty$ as $n \to \infty$.

(ii) Assume that $n^{\frac{1}{2}} h_{n,\bar{T}} \to \infty$.

In section 2.2, we derive the consistent estimator of the exploding diffusion coefficient (2.10) for SDE of the kind (2.6).

2.2 Estimator

Let X_t be a solution of the SDE in Eq. (2.6) over the time interval $[0,\bar{T}]$. The processX_t is recorded discretely at n equally spaced time point over a time interval $[0,\bar{T}]$.[2] Let subdivide the time interval $[0,\bar{T}]$ in n steps equal length,e.i. $\{t_1 = \Delta_{n,\bar{T}}, t_2 = 2\Delta_{n,\bar{T}}, \ldots, t_n = n\Delta_{n,\bar{T}}\}$, and hence we obtain$X_t = X_{\Delta_{n,\bar{T}}}, X_{2\Delta_{n,\bar{T}}}, \ldots, X_{n\Delta_{n,\bar{T}}}$ where $\Delta_{n,\bar{T}} = \frac{\bar{T}}{n}$. Here, we consider that the number of sampled point (n) is large enough, and the period $\bar{T}$ is fixed. Thus, we will explore the limit theory of the proposed estimator as $n \to \infty$, $\bar{T} < \infty$ and $\Delta_{n,\bar{T}} \to 0$. From assumption 3, we knew that X_t is a continuous semimartingale and thus its local time can be defined as (Karatzas and Shreve 2012):

$$L_X(t,x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t I_{[x,x+\varepsilon)}(X_s) d[X]s, \tag{2.11}$$

where $I(\cdot)$ is an indicator function and $[X]_t$ is the quadratic variation process of X_t. The term $L_X(t,x)$ embodies the amount of time the process X spends in the neighborhood of the point x. The time measures in units of the quadratic variation process $([X]_s)$ (see Bandi 2002). Let us now consider the diffusion process of X with the quadratic variation $\sigma^2(X_s)$ and thus (2.11) reduces to

$$L_X(t,x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t I_{[x,x+\varepsilon)} \sigma^2(X_s) ds. \tag{2.12}$$

[2]The assumption that the observed prices X_{t_i} are equally spaced is not too restrictive, in fact, we can generalize the results below to the case the observations X_{t_i} are not equally spaced.

If we divide the both side of the Eq.(2.13) by $\sigma^2(x)$ and obtain:

$$\bar{L}_X(t,x) = \frac{L_X(t,x)}{\sigma^2(x)} = \lim_{\varepsilon\to 0}\frac{1}{\sigma^2(x)}\frac{1}{\varepsilon}\int_0^t I_{[x,x+\varepsilon)}\sigma^2(X_s)ds. \tag{2.13}$$

The process $\bar{L}_X(t,x)$ calls "chronological local time" and records the amount of time spends by a process X at the neighborhood of a point x. According toBandi and Phillips (2009), by using the sampling scheme above, we can approximate the chronological-local-time of the diffusion process as:

$$L_X^n(t,x) = \frac{\bar{T}}{nh_{n,\bar{T}}}\sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_{n,\bar{T}}}\right). \tag{2.14}$$

The following result is obtained by Florens-Zmirou (1993)

Proposition 2 (Florens-Zmirou (1993))**.**

Assume that the assumption (3) *holds. Suppose* $n^4 h_{n,\bar{T}} \to 0$ *as* $n \to \infty$ *(with fixed time* $\bar{T}$*), then* $L_X^n(t,x) \to \bar{L}_X(t,x)$ *in* $\mathscr{L}^2$ *sense.*

The nonparametric estimator of diffusion process proposed by Florens-Zmirou (1993) is of the class (2.1). The Florens-Zmirou estimator is based on the following quantity

$$V_t^n = \frac{1}{h_{n,\bar{T}}}\sum_{i=1}^{n-1} K\left(\frac{X_i - x}{h_{n,\bar{T}}}\right)(X_{(i+1)\Delta_{n,\bar{T}}} - X_{t_i})^2. \tag{2.15}$$

Dividing V_t^n by L_t^n to obtain the consistent estimator of diffusion process proposed by Florens-Zmirou (1993):

$$S_{FZ}^n(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_{n,\bar{T}}} - x}{h_{n,\bar{T}}}\right)\bar{\sigma}_{FZ}^2(X_{i\Delta_{n,\bar{T}}})}{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_{n,\bar{T}}} - x}{h_{n,\bar{T}}}\right)} \quad \text{where} \tag{2.16}$$

$$\bar{\sigma}_{FZ}(X_t) = \frac{(X_{t+\Delta_{n,\bar{T}}} - X_t)^2}{\Delta_{n,\bar{T}}}, \tag{2.17}$$

The following result is also obtained by Florens-Zmirou (1993)

Proposition 3 (Florens-Zmirou (1993))**.**

Suppose $n \to \infty$ *and we have* $nh_{n,\bar{T}}^4 \to 0$*, then* $V_t^n \to \sigma^2 L_t$ *in the* $\mathscr{L}^2$ *sense. The convergence is almost sure if* $\log n/nh_{n,\bar{T}}^2 \to 0$.

Combining the results of the Propositions 2 and 3, thus, we see that by dividing V_t^n by L_t^n the consistent estimator Florens-Zmirou of $\sigma(\cdot)$ is derived.

In Chapter 1, we underlined that during the equity flash crash, the volatility spikes (see Andersen et al. 2016 and Kirilenko et al. 2017 for some details). If we use the Florens-Zmirou estimator in Eq. (2.16) or one of the above-mentioned non-parametric estimators to estimate the exploding diffusion coefficient, and the estimator may fail to estimates more accurately the diffusion coefficient since the volatility obtained with a spot volatility $\bar{\sigma}_{FZ}(\cdot)$ is higher during an extreme event. It is, therefore, reasonable to propose the non-parametric estimator to estimate more accurately the diffusion coefficient during an extreme event.

2.2.1 Correction of Florens-Zmirou estimator

Suppose that X_t is a solution of SDE (2.6) and applying the Euler discretization rule to obtain

$$X_{(i+1)\Delta_{n,\bar{T}}} - X_{i\Delta_{n,\bar{T}}} = \mu(X_{(i)\Delta_{n,\bar{T}}})\Delta_{n,\bar{T}} + \sigma(X_{(i)\Delta_{n,\bar{T}}})(W_{(i+1)\Delta_{n,\bar{T}}} - W_{i\Delta_{n,\bar{T}}}). \tag{2.18}$$

We can also write (2.18) in the following way

$$(X_{(i+1)\Delta_{n,\bar{T}}} - X_{i\Delta_{n,\bar{T}}} - \mu(X_{(i)\Delta_{n,\bar{T}}})\Delta)^2 = \sigma^2(X_{(i)\Delta_{n,\bar{T}}})(W_{(i+1)\Delta_{n,\bar{T}}} - W_{i\Delta_{n,\bar{T}}})^2. \tag{2.19}$$

Set $t_i = i\Delta_{n,\bar{T}}$ and by taking the expectation of the above equation and then (2.19) becomes

$$\sigma^2(X_{t_i}) = \Delta_{n,\bar{T}}^{-1} E[(X_{t_{i+1}} - X_{t_i} - \mu(X_{t_i})\Delta_{n,\bar{T}})^2], \tag{2.20}$$

where $\mu(x)$ is a suitable drift estimator. Using the fact that the drift coefficient explodes at a magnitude higher than the one of a diffusion coefficient, and therefore by including the drift coefficient $\mu(\cdot)$ in the spot volatility estimatorin Eq. (2.20), then, the estimator (2.20) will provide a smaller volatility estimation during the flash crash. Thus, spot volatility estimator (2.20) provides a more accurate estimation of the volatility than $\Delta_{n,\bar{T}}^{-1}(X_{t_{i+1}} - X_{t_i})^2$

during the flash crash. Moreover, the drift coefficient $\mu(\cdot)$ is the function of an unknown function g. If $g.$ is known, then we can rely on the parametric estimator of $\mu(x)$ to use the spot volatility estimator (2.20). If $g''(x)$ converges to zero for every x, then a proposed spot volatility estimator (2.20) coincides to the estimator (2.17). However, the function g is unknown, and we may not rely on the semi-parametric estimator of $\mu(x)$ to estimate a drift coefficient $\mu(x)$. Thus, it may be impossible to use the spot volatility estimator in Eq. (2.17). On the other hand, we can also rely on the nonparametric estimator of a drift coefficient $\mu(x)$. Let $\hat{\mu}(\cdot)$ be a nonparametric estimator of$\mu(x)$ and defines as:

$$\hat{\mu}(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_{n,\bar{T}}} - x}{h_{n,\bar{T}}}\right)(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_{n,\bar{T}}} - x}{h_{n,\bar{T}}}\right)\Delta_{n,\bar{T}}}. \tag{2.21}$$

The estimator in Eq. (2.21)) is a similar version of a nonparametric estimator for a drift coefficient proposed by Bandi and Phillips (2003). Under some conditions (including $t_i \to \infty$), Bandi and Phillips (2003) showed that the estimator (2.21) is a consistent estimator of a drift coefficient and its asymptotic distribution is normal. However, if the period t_i is fixed as in our case, then the drift coefficient cannot be identified and it would diverge at a rate given by the square root of h_n (see Theorem 2.1 in Bandi 2002). Moreover, we can show that $\Delta_{n,\bar{T}}\hat{\mu}(x)$ does not that diverge (see Lemma 1).

Using (2.21) and we then rewrite the above spot volatility estimator as follows:

$$\sigma^2(X_{t_i}) = \Delta_{n,\bar{T}}^{-1}\mathbb{E}[(X_{t_{i+1}} - X_{t_i} - \hat{\mu}(X_{t_i})\Delta_{n,\bar{T}})^2]. \tag{2.22}$$

In general, the spot volatility estimator (2.22) has a smaller error than (2.17). Using the Eq.(2.21) and thus we introduce the correction of the Florens-Zmirou estimator as follows,

$$S^n_{CFZ}(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_{n,\bar{T}}} - x}{h_{n,\bar{T}}}\right)(X_{t_{i+1}} - X_{t_i} - \hat{\mu}(X_{t_i})\Delta_{n,\bar{T}})^2}{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_{n,\bar{T}}} - x}{h_{n,\bar{T}}}\right)\Delta_{n,\bar{T}}}. \tag{2.23}$$

The proposed estimator (2.23) calls "correction of Florens-Zmirou estimator".

Consistency and asymptotic theory of the proposed Estimator

We now turn attention to the consistency and asymptotic theory of the proposed diffusion estimator (2.21). We first discuss the following Lemma.

Lemma 1. *Assume the assumptions* (3), (4) *and* ((i)) *hold. If* $n^4 h_{n,\bar{T}} \to \infty$ *and* $nh_{n,\bar{T}}^{1/2} \to \infty$, *then we obtain*

(i)

$$\hat{\mu}(X_{t_i}) = \frac{1}{\sqrt{h_{n,\bar{T}}}} O_p\Big(1\Big) \tag{2.24}$$

(ii)

$$\frac{\Delta_{n,\bar{T}}}{h_{n,\bar{T}}} \sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_{n,\bar{T}}} - x}{h_{n,\bar{T}}}\right) \xrightarrow{\mathscr{L}^2} L_X(\bar{T},x). \tag{2.25}$$

Proof. The proof is given in Appendix (A.1) □

Remark 8. *Note that* $\mathscr{L}^2$ *in the Lemma 2.25 means the convergence in mean square error. The condition 2.25 is similar to a Proposition* 2 *in Florens-Zmirou (1993) and Proposition* (2). *In the Lemma* (1) *we note that* $\hat{\mu}(X_{t_i})$ *explodes, but* $\Delta_{n,\bar{T}}\hat{\mu}(X_{t_i})$ *does not so, we can use it in* (2.23).

In the following Proposition we show the consistency of diffusion estimator (2.23) holds.

Proposition 4 (Consistency of S^n_{CFZ})**.**

Assume the assumptions (4), (4) *and* ((i)) *hold. If* $nh_{\bar{T}}^4 \to 0$ *as* $n \to \infty$, *then* $S^n_{CFZ}(x)$ *is consistent estimator of* $\sigma^2(x)$ *in the* $\mathscr{L}^2$ *sense.*

Proof. The proof is given in the appendix A.2. □

Proposition 5 (Asymptotic theory of S^n_{CFZ})**.** *Suppose that Assumptions* (3), (4) *and* (5) *hold. If* $n \to \infty$ *(with fixed* $\bar{T}$*),* $\Delta_{\bar{T},n} \to 0$ *and* $h_n \to 0$, *then*

$$\sqrt{nh_{n,\bar{T}}}\left(\frac{S^n_{CFZ}(x)}{M^2(x)} - 1\right) \xrightarrow{\mathscr{L}} \frac{1}{\sqrt{\bar{L}_X(\bar{T},x)}} \mathscr{N}(0,1) \quad \forall x \in \mathfrak{D}. \tag{2.26}$$

The above convergence is in law and $\mathcal{N}(0,1)$ is a standard normal variable.

Proof. The proof is given in the appendix (A.3). □

2.3 Simulation Results

In this section, we test the unbiased nature of the proposed estimator on the simulate data. We also compare our proposed estimator (2.23) to a classical Florens-Zmirou estimator (2.16) by mean of Monte Carlo simulation in-small sample. We simulate discrete sample data using two types of models. We first simulate the sample data using the CEV model discussed below. With the CEV model, we simulate sample data which are different from the flash crash prices. The idea is to asses the unbiased nature of the proposed estimator when estimating a non-exploding diffusion coefficient. Second, we consider models (3), (5) introduced in Chapter 1. With these models, we simulate the flash crash price type of data. Then, we use our estimator and Florens-Zmirou estimator to estimate the exploding diffusion coefficient ($\sigma(X)^2$) on these simulate data.

Let $X_{i\Delta_{n,\bar{T}}}$ denote the simulate discrete data, $\Delta_{n,\bar{T}} = \frac{\bar{T}}{n}$ and $i = 1,\ldots,n$. In this work, we consider a Gaussian kernel, e.i.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}. \tag{2.27}$$

The bandwidth is chosen as $h_{\Delta_{n,\bar{T}}} = h_s.\bar{s}.n^{-1/5}$ where h_s is set to be equal to 1.06 while $\bar{s}$ is a standard deviation.

Experiment I: Simulation of the CEV model

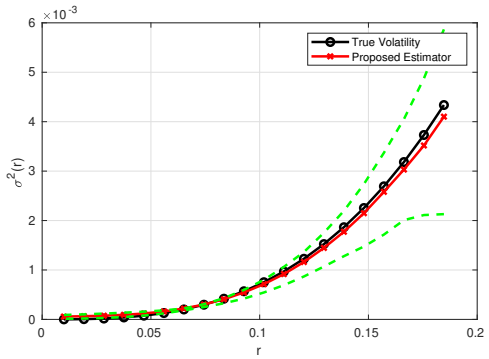
We simulate a small sample of discrete data using the parametric model proposed by Chan et al. (1992) which includes some popular models, such as CIR (1995), Vasicek (1977) and CEV model to asses the performance of the proposed estimator in Eq (2.23). The model under consideration is the following

$$dr_t = \kappa(\theta - r_t)dt + \varepsilon r_t^{\gamma} dW_t, \quad t \in [0,\bar{T}], \tag{2.28}$$

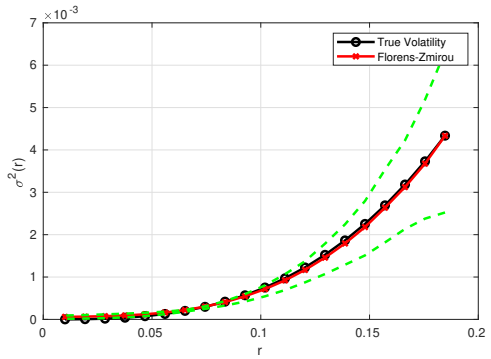
where W is standard Brownian motion and $\kappa, \theta, \varepsilon, \gamma$ are constants. A parameter γ measures the degree of dependence of the variance of the interest rate. We simulate the model (2.28) using the first-order Euler discretization scheme to obtain the monthly trajectories of the short rate r, and the time step between two consecutive observation is equal $\frac{1}{12}$. In the literature, many authors have estimated the parameters of the model in Eq. (2.28), and in this work, we consider the CEV parameters estimated by Jiang (1998) using indirect interference. Thus, the estimated parameters under consideration are the following $\hat{\theta} = 0.079, \hat{\kappa} = 0.093, \hat{\gamma} = 1.474, \hat{\varepsilon} = 0.794$. We simulate 5000 sample paths of the interest rate with the time step between observing prices is $\Delta = \frac{1}{12}$. We compare the performance of proposed estimator (2.23) to the popular one of Florens-Zmirou (1993). We simulate 1000 trajectories of length 5000 with the model in Eq. (2.28) and estimate the volatility using the non-parametric estimator in Eq. (2.23) and Florens-Zimrous estimator in Eq. (2.16).

Figure (2.1) plots the average diffusion coefficient estimated with the Florens-Zmirou estimator in Eq. (2.16) and the newly proposed on the 1000 trajectories of length 5000 of short term interest rate generated by CEV model in Eq. (2.28).

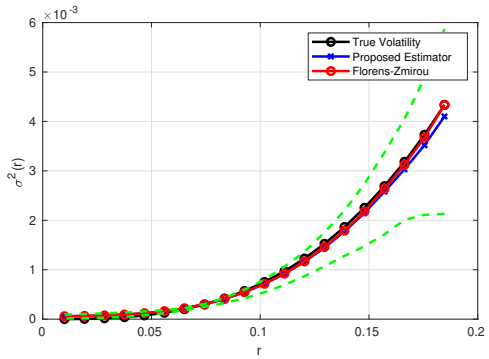
Figure 2.1 The generated Monte Carlo squared diffusion coefficient according to the CEV model; e.i. $\sigma^2 = \hat{\sigma}^2 r^{2\gamma}$ (solid tick line) together with the average estimate, using 1000 replications, with two methods: right panel 2.1(b) is the proposed estimator (2.23) and the left panel 2.1(a) the classical estimator (2.16). In a center panel 2.1(c) we combine the newly proposed estimator 2.23 and Florens-Zmirou (2.16). The dash green lines are 95% and 5% confidence interval draw from the simulations. We observe all estimators are unbiased. Correction is not of help here



(a)



(b)



(c)

From Figure (2.1), we first observe all the estimators are unbiased. Figure 2.1(b) shows that the volatility estimated with the newly proposed non-parametric estimator is similar to true one up to 15%, while slightly deviates the true one above 15%. The higher value of interest rate corresponds to the higher volatility while the lower one corresponds to the lower volatility. We also notice that the 95% confident band is narrower in the begin while it tends to become wilder close to the end. It is maybe due to the lack of enough data towards the end. In Figure 2.1(a), we observe that the volatility obtained with the Floren1s-Zmirou estimator is similar to the true one even above 15%. Similarly, as in Figure 2.1(b), the higher interest rate corresponds to the higher

estimate volatility while its lower values correspond to the lower volatility. The 95% confident band are also narrower in the begin while it tends to become wilder close to the end.

From Figure 2.1(c), we confront the two estimators; newly proposed estimator and the Florens-Zmirou estimator. Figure 2.1(c) still confirms there is almost no substantial difference between the two estimators as the estimated volatility obtained with the proposed non-parameter estimator and Florens-Zmirou estimator are well between 95%confident band. Thus, the correction is not of help here.

Experiment II: Flash crash data simulate with exploding model

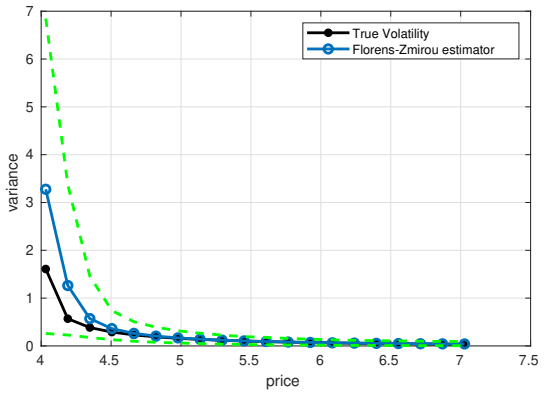
The purpose of this subsection to asses the performance of our propose estimator using the simulate flash crash data. We test the bias nature of the propose estimator and Florens-Zmirou estimator on the simulate flash crash prices. We consider the parametric models (2.3) and (2.4) to simulate the sample of the flash crash prices. To generate the trajectories of the exploding models in Eq. (2.4), we use the dynamic Euler algorithm introduced in Section 1.4. To use the dynamic Euler algorithm, we must set the following parameters minRMSE, $\alpha, \beta, c_\sigma, c_\mu, N$; t_{end}. Here we set minRMSE $= 5 \times 10^{-7}, \alpha = 0.995, \beta = 0.25, c_\sigma = 0.01, c_\mu = 0.7, N = 1000$; $t_{end} = 1 - 10^{-8}$. The parameters of the models (2.3) and (2.4) are estimated from the E-mini S&P 500 flash crash (SPY) of May 2010 (see Section 1.3 of Chapter 1). The estimated parameters of the models in Eq. (2.4) and (2.3) are expressed in annually form; $k = 1.856764; a_1 = -0.365213$; and $\sigma_1^* = 0.265886$, and $k = 5.583505, a = 0.313882$ and $\sigma^* = 0.278899$. We also assume that the starting price X_0 and the price before a flash crash $\bar{X}$ are equal to 5.5, e.i. $X_0 = \bar{X} = 5.5$. We assume that the exploding price x_0 to be equal to 4. To compare the performance of a proposed estimator in Eq. (2.23) to the Florens-Zmirou estimator using the simulate flash crash data, we simulate 1000 replicate trajectories of the exploding models (2.4). We then compute the proposed estimator (2.23) and Florens-Zmirou estimator (2.16) on these simulations.

In Figure (2.2), we plot the average diffusion coefficient estimated with the Florens-Zmirou estimator 2.16 and our proposed estimator on the 1000 trajectories of flash crash prices generate with the exploding model (2.4). Figure 2.2(a) shows that the volatility obtained with the Florens-Zmirou estimator is similar to the true one up to 4.34, whereas suddenly increases and deviates from the true one above 4.34. Furthermore, the estimated volatility obtained with Florens-Zmirou estimator is almost two times higher than the true one in the exploding region $(4.34 - 4)$. This result may indicate that the Florens-Zmirou estimator is not able to estimate more accurately the exploding diffusion coefficient. In contrast to the scenario observed in the previous case, the result

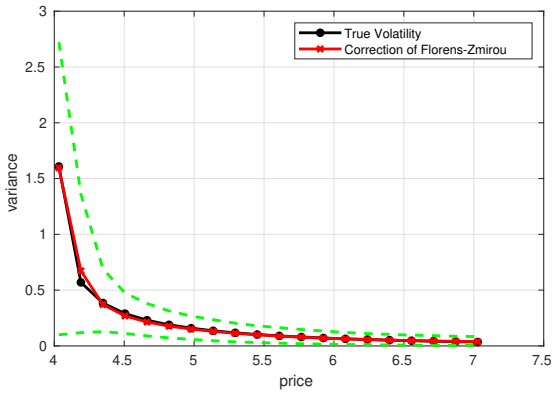
of Figure 2.2(b) shows the average volatility obtained with the proposed non-parametric estimator is similar to true one even up to 4.34. This result may indicate, including the non-parametric estimator of drift coefficient in spot volatility estimator allows us to estimate more accurately the exploding diffusion coefficient. Because it is in this region (4.34-4), where the drift coefficient $|\mu(x)|$ explodes with a magnitude higher than the one of diffusion coefficient, that our estimator shows a significant improvement over the Florens-Zmirou estimator when estimating the exploding diffusion coefficient.

From Figure 2.2(c), we confront the two estimators; proposed estimator and the Florens-Zmirou estimator. Figure 2.2(c) still confirms there is a substantial difference between the two estimators. The estimated volatility obtained with the proposed non-parameter estimator is well between 95% confident bands, whereas the one obtained with Florens-Zimrou falls outside 95% confident band in the region of $(4.5-4)$. Thus, the correction is of help here.

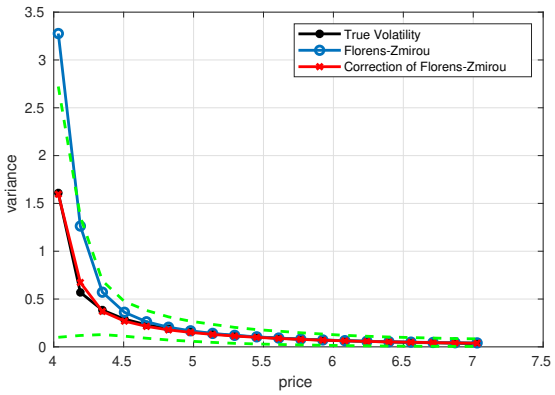
Figure 2.2 The generated Monte Carlo squared exploding diffusion coefficient according to the exploding model 2.4; e.i. $\sigma^2(x) = \sigma_1^{2*}\left(\frac{k-\ln(\bar{x}-x_0)}{k-\ln(x-x_0)}\right)^{2a/\sigma_1^*}$ (solid tick line) together with the average estimate, using 1000 replications, with two methods: right panel 2.2(a) is the proposed estimator (2.23) and the left panel 2.2(b) the classical estimator(2.16). In a center panel (2.2(c)) we compare the estimates obtained with both estimator. Dash lines are 95% and 5% confidence interval fort the proposed estimator draw from the simulations. We observe the propose estimators is unbiased while the Florens-Zmirou estimator is biased around the exploding price. Correction is of help here.



(a)



(b)

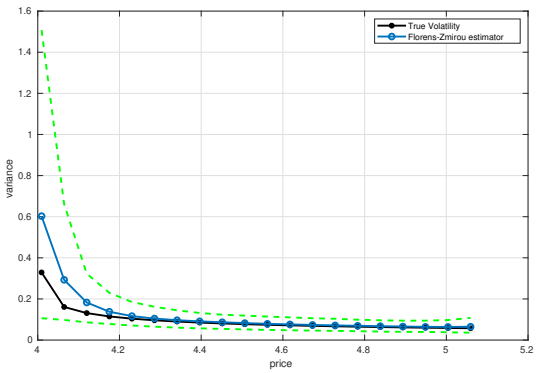


(c)

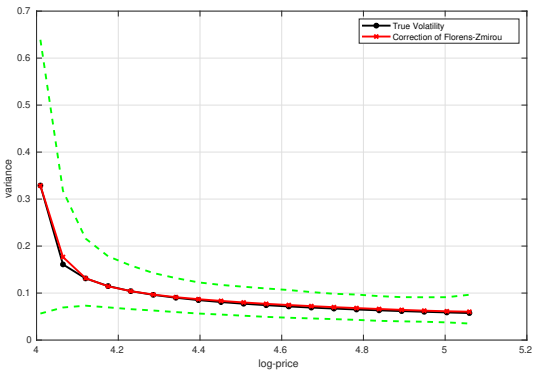
We now turn our attention to the results obtained flash crash data simulated with the model in Eq. (2.3). Figure (2.5) plots the average diffusion coefficient estimated with the Florens-Zmirou estimator and our proposed estimator on the 1000 trajectories of flash crash prices simulates by the exploding model (2.3). In Figure 2.5, we observe the exact scenario as in the above case. We also see that the proposed estimator shows

significant improvement over the FZ estimator. We conclude that when $|\mu(x)|$ is higher and, the inclusion of the nonparametric estimator of drift coefficient in the spot volatility estimator allow us to improve the accuracy of the estimator in the estimation of exploding diffusion coefficient.

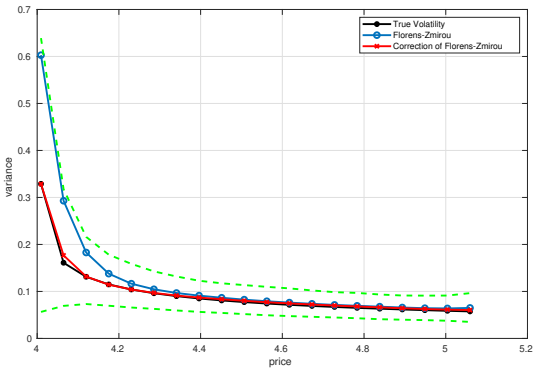
Figure 2.3 The generated Monte Carlo squared exploding diffusion coefficient according to the exploding model 2.4; e.i. $\sigma^2(X) = \frac{\sigma^{*2}(k-a\ln(X_t-x_0))^{\frac{2}{\sigma^*}}}{(k-a\log(\bar{X}-x_0)^{\frac{2}{\sigma^*}}}$ (solid tick line) together with the average estimate, using 1000 replications, with two methods: right panel (2.2(b)) is the proposed estimator (2.23) and the left panel (2.2(a)) the classical estimator(2.16). In a center panel (2.2(c)) we compare the estimates obtained with both estimator. Dash lines are 97.5% and 2.5% confidence interval fort the proposed estimator draw from the simulations. We observe the propose estimators is unbiased while the Florens-Zmirou estimator is biased around the exploding price. Correction is of help here.



(a)



(b)



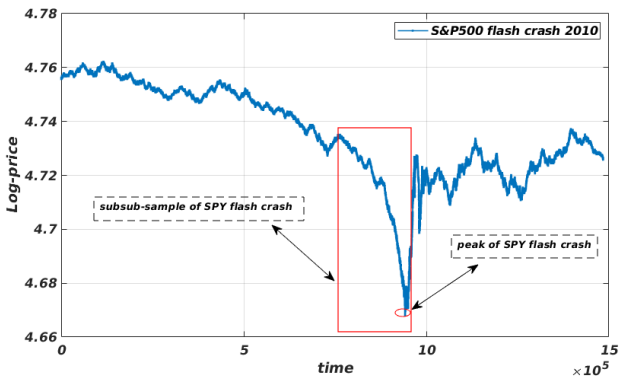
(c)

2.4 Application to flash crash data

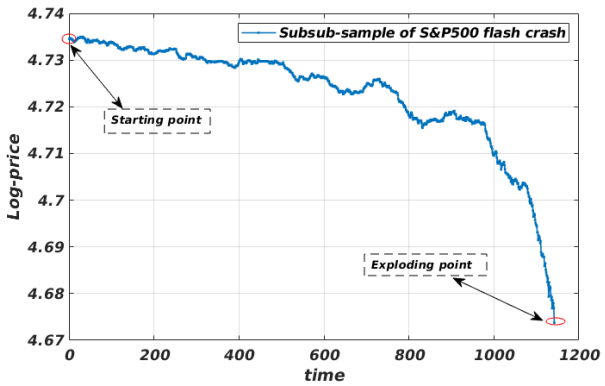
The purpose of this section is to asses the performance of proposed estimator and Florens-Zmirou (FZ) estimator when applying to the E-mini S&P 500 flash crash (SPY) of May, 2010. We estimate the diffusion coefficient $\sigma^2(.)$ of the exploding model (2.6) using the SPY flash crash data. In Figures 2.4(a) and 2.4(b) we plot the log-price of SPY flash crash and its sub-sample. Figure 2.4(a) shows on a red-point the peaks of the SPY flash crash, 4.6738. Based on the theory of Section 2.1.1, we consider the point 4.6738 of SPY flash crash as the exploding price of SPY flash crash. To correctly implement the proposed estimator on the SPY flash crash data, we must be consistent with the theory of Section 2.1.1. Thus, we cannot consider the full sample of SPY flash crash prices, since we are only interested in the region where the prices explode the most.

In this section, we implement the proposed estimator and Florens-Zmirou estimator on the subsub-sample of SPY flash crash (see Figure 2.4(b)). The subsub-sample of SPY flash crash data starts point at 4.7350 and stops at 4.6738 as we can see from the Figure 2.4(b). The total number of observations of the subsub-sample of SPY flash crash data is 1142.

Figure 2.4 Figure 2.4(a) shows the log-price of S&P500 flash crash data on 06/05/2010 and Figure 2.4(b) shows its subsub-sample.



(a)



(b)

To implement the proposed estimator and FZ estimator on the sub-sub-sample of SPY flash crash price, we use a large h_s compared to that one used on simulations. We use $h_s = 4$. For example, Renò (2008) and Stanton

(1997) also used the same value of h_s when estimating the continuous-time models for the interest rate. We set the time step $\Delta = \frac{1}{252\times 23400}$ so that the estimate volatility is expressed in annualized form.

Let E denotes the difference between the volatility estimated with FZ estimator and the proposed estimator (error), e.i.

$$E(x) = S^n_{FZ}(x) - S^n_{CFZ}(x), \tag{2.29}$$

where $S^n_{FZ}(.)$ is a FZ estimator (2.16) and $S^n_{CFZ}(.)$ the proposed estimator (2.23).

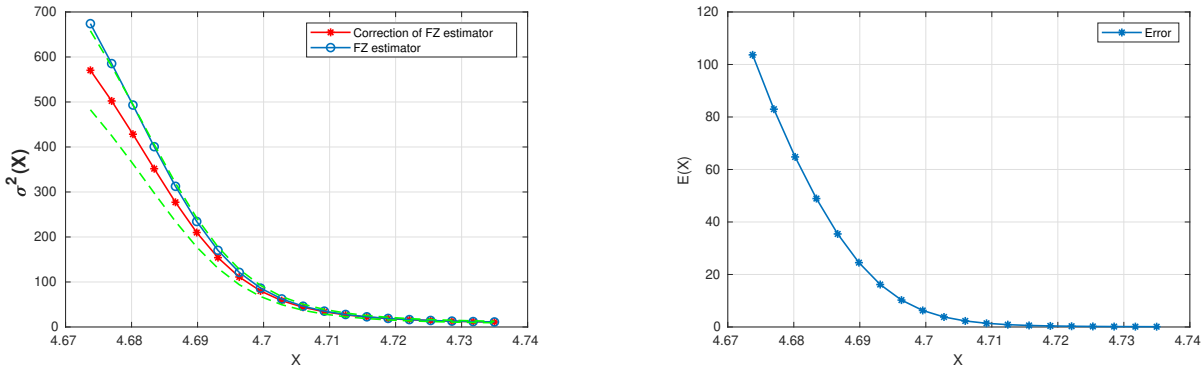
In Figures 2.5(a), we compare volatility $\sigma^2(X)$ estimates with our proposed estimator in Eq 2.23 to that estimates with the classical FZ estimator 2.16. Figure 2.5(b) plots the error computing using Eq. (2.29). Figure 2.5(a) shows the volatility $\sigma^2(X)$ estimates with FZ estimator is similar to that estimated with our estimator up to 4.7, however, above 4.7, the FZ estimator estimates higher volatility compared to the volatility estimates with our estimator. As also shown in Figure 2.16 that the error is significantly higher (from 6.2962 up to 103.6628) in the interval of $(4.7 - 4.6738)$. Furthermore, the volatility $(\sigma^2(X))$ estimates with FZ estimator falls outside the 95% confidence level in the interval of $(4.7 - 4.6738)$. These results suggest that the new estimator appears to estimate more accurately volatility $\sigma^2(X)$ compared to the classical FZ estimator during the extreme event such as a flash crash.

2.5 Conclusion and possible extensions

In this work, we proposed a new nonparametric estimator for estimating the diffusion function of the exploding model. The proposed estimator is constructed based on the theory of Nadaraya-Watson estimator. The proposed estimator is similar to the FZ estimator. However, in our case, we include the nonparametric estimator of the drift coefficient in the spot volatility estimator that helps us to estimate more accurately an exploding diffusion coefficient. Our estimator is consistent and asymptotically normally distributed as the discretization step (Δ) goes to zero.

We test unbiased nature of our proposed estimator on two types of the simulate data, namely the flash crash prices type and non-flash crash prices type. We compared the proposed estimator to the classical Florens-Zmirou

Figure 2.5 Figure 2.5(a) plots estimation of the exploding diffusion coefficient $\sigma^2(X)$ on subsub-sample of SPY flash crash (2010) data set. Red star line: the estimate obtained with the proposed estimator. Dashed lines are 5% and 95% confidence intervals for the proposed estimator, computed using 2.26. Dotted blue line: the estimate obtained with the FZ estimator. Figure 2.5(b) plots the bias error between the estimate obtained with the FZ estimator and the proposed estimator, computed using 2.29.



(a) (b)

by mean of Monte Carlo simulations. We simulate the flash crash prices using the exploding models (2.4) and (2.3) and the non-flash crash prices using the CEV model. The Monte Carlo showed that both estimators estimated more accurately the diffusion coEfficients $\sigma(X)^2$ of CEV model. However, in the case of exploding models (2.4) and (2.3), our proposed estimator showed a significant improvement over FZ estimator, as the latter estimated more accurately the diffusion coefficient $\sigma(X)^2$ of models(2.4) and (2.3) compared to FZ estimator. Lastly, we calculate volatility ($\sigma(X)^2$)of the SPY flash crash (2010). Lastly, we calculate volatility $\sigma(X)^2$ on SPY flash crash (2010). The results showed that the estimated volatility obtained with both estimators are similar in the interval of $4.7 - 4.6738$, however, above 4.7, the volatility estimated with our estimator is lower compared to the one estimated with the FZ estimator. The result also showed that the volatility obtained with FZ estimator falls outside the 95% confidence level. Using our proposed estimator, we can reliably estimate the volatility during the extreme event such as a flash crash.

On the other hand, some improvement can be considered. One can consider a different type of the drift estimator instead of $\hat{\mu}(x)$. For example, we can use the drift estimator of Bandi (2002) in the spot volatility estimator (2.22). We note that the choice of the bandwidth h_s can affect the estimation of the exploding volatility. In this work we did not cover this issue, but some work can still be done in order to select the optimal h_s.

A Appendix

A.1 Proof of Lemma 1

We only prove (2.24). Observe that

$$\Delta_{n,\bar{T}}\hat{\mu}(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h_{n,\bar{T}}}\right)\left(X_{t_{i+1}}-X_{t_i})\Delta\right)}{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_{n,\bar{T}}}-x}{h_{n,\bar{T}}}\right)}.$$

We will not do go into detail for the sketch of the proof of condition (2.24) since it is exactly similar to the proof of Theorem 2.1 in Bandi (2002) (see pg:101-106). Bandi (2002) proved that under a condition $\bar{T}$ is fixed, the drift estimator $\hat{\mu}(x)$ diverges at rate given by the square root of $h_{n,\bar{T}}$ as $n \to \infty$, i.e.

$$\sqrt{h_{n,\bar{T}}L_x(t,x)}(\hat{\mu}(x)-\mu(x)) \Rightarrow N(0,\sigma^2(x)). \tag{A.1}$$

With Eq.(A.1) gives

$$\sqrt{h_{n,\bar{T}}}(\hat{\mu}(x)) \Rightarrow (L_x(t,x))^{-1/2}N(0,1/2\sigma^2(x)), \tag{A.2}$$

and

$$\hat{\mu}(x) = \frac{1}{\sqrt{h_{n,\bar{T}}}}O_p\Big(1\Big). \tag{A.3}$$

Observe from Eq.(A.3) if the drift estimator $\hat{\mu}(x)$ diverges at rate given by $\sqrt{h_{n,\bar{T}}}$ and then by multiplying (A.3) by $\Delta_{n,\bar{T}}$, and $\Delta_{n,\bar{T}}\mu(x)$ becomes

$$\Delta_{n,\bar{T}}\hat{\mu}(x) = \frac{\Delta_{n,\bar{T}}}{\sqrt{h_{n,\bar{T}}}}O_p\Big(1\Big). \tag{A.4}$$

Then, one can see $\Delta_{n,\bar{T}}\hat{\mu}(x)$ converges eventually to zero at rate given by $\frac{\sqrt{h_{n,\bar{T}}}}{\Delta_{n,\bar{T}}}$ as $n \to \infty$ since $\frac{\Delta_{n,\bar{T}}}{\sqrt{h_{n,\bar{T}}}} \xrightarrow{n\to\infty} 0$. Note that if $\Delta_{n,\bar{T}}\hat{\mu}(x)$ converges to zero does not guarantee that $\Delta_{n,\bar{T}}\hat{\mu}(X_{t_i})$ might converge to zero.[3] Now, we need to prove the following result

$$\Delta_{n,\bar{T}}\hat{\mu}(X_{t_i}) = \frac{\Delta_{n,\bar{T}}}{\sqrt{h_{n,\bar{T}}}} O_p\Big(1\Big). \tag{A.5}$$

Before proving a result (A.5), we need the recall some results in Nicolau (2003b). Nicolau (2003b) proved that under the condition $T \to \infty$, if $\hat{\mu}(x) \xrightarrow{p} \mu(x)$ then $\hat{\mu}(X_{t_i}) \xrightarrow{p} \mu(X_{t_i})$ (see the proofs of the Theorems $4-6$ pg:$20-36$ in Nicolau (2003b)). In the same spirit for the proof of the Theorems $4-6$ in Nicolau (2003b), as $\Delta_{n,\bar{T}}\hat{\mu}(x) \xrightarrow{p} 0$ then one can also show that $\Delta_{n,\bar{T}}\hat{\mu}(X_{t_i}) \xrightarrow{p} 0$, e.i.

$$\Delta_{n,\bar{T}}\hat{\mu}(X_{t_i}) = \frac{\Delta_{n,\bar{T}}}{\sqrt{h_{n,\bar{T}}}} O_p\Big(1\Big). \tag{A.6}$$

The proof of the (2.25) can be reached in Florens-Zmirou (1993), see the proof of Proposition 2 with the kernel given as am indicator function.

A.2 Proof of Proposition 4

We want to show that $S^n_{new}(x) \xrightarrow{p} \sigma^2(x)$ **as** $n \to \infty$. Let $(\Omega', (\mathscr{F}'_t, \mathscr{P}')_{(0\le t\le \bar{T})}, \mathscr{F})$ be an auxiliary probability space, and let denote by $\mathbb{E}'$ the expected value in Ω' and $\mathbb{E}$ the expected value in Ω. For all $\omega \in \Omega$, consider a solution $X_t(\omega)$ of Eq. (2.6) and define the process $Z_t(\omega)$ as

$$dZ_t(\omega) = \sigma(X_t(\omega))dW'_t, \tag{A.7}$$

where W'_t is a standard Brownian motion in $(\Omega', (\mathscr{F}'_t,)_{(0\le t\le \bar{T})}, \mathscr{P}')$ and $\sigma(X_t(\omega))$ is the realization of $\sigma(X_t)$. From assumption (3), one can guarantee the existence and uniqueness of the solution of (A.7). Then one can construct the

[3] see Davidson (1994), Chapiter 21

random variable in Ω by taking the expectation Ω'. In the remain of the proofs we fix $\Delta_{n,\bar{T}} = \Delta$ and $h_{n,\bar{T}} = h$. Observe that

$$\begin{aligned}\frac{1}{\Delta}\mathbb{E}'[(Z_{i+1}-Z_i)^2] = \frac{1}{\Delta}\int_{t_i}^{t_{i+1}} \sigma^2(X_s(\omega))ds \quad \text{with} \quad Z_{i+1} = Z_{t_{i+1}}(\omega)\\ \Rightarrow \mathbb{E}[\frac{1}{\Delta}\mathbb{E}'[(Z_{i+1}-Z_i)^2]|\mathscr{F}_{t_i}] = \frac{1}{\Delta}\int_{t_i}^{t_{i+1}} \mathbb{E}[\sigma^2(X_s(\omega))|\mathscr{F}_{t_i}]ds. \end{aligned} \tag{A.8}$$

Using the equation (2.6) we ca show that

$$\frac{1}{\Delta}\mathbb{E}[(X_{t_{i+1}}-X_{t_i})^2|\mathscr{F}_{t_i}] = \frac{1}{\Delta}\int_{t_i}^{t_{i+1}} \mathbb{E}[\sigma^2(X_s)|\mathscr{F}_{t_i}]ds + \frac{1}{\Delta}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} \mu(X_s)ds\right)^2|\mathscr{F}_{t_i}\right], \tag{A.9}$$

Since (A.8) and (A.9) depend on $\int_{t_i}^{t_{i+1}} \mathbb{E}[\sigma^2(X_s)|\mathscr{F}_{t_i}]ds$, then we can easily prove that almost surely,

$$\mathbb{E}[\mathbb{E}'[(Z_{i+1}-Z_i)^2]|\mathscr{F}_{t_i}] = \mathbb{E}[(X_{t_{i+1}}-X_{t_i})^2|\mathscr{F}_{t_i}] - \mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} \mu(X_s)ds\right)^2|\mathscr{F}_{t_i}\right], \tag{A.10}$$

where X_t is a solution of (2.6). Now, using the fact that $\mathbb{E}[(X_{t_{i+1}}-X_{t_i})^4|\mathscr{F}_{t_i}] < \infty$ and by applying the Cauchy-Schwartz inequality, one can prove almost surely that

$$\mathbb{E}[\mathbb{E}'^2[(Z_{i+1}-Z_i)^2]|\mathscr{F}_{t_i}] \leq \mathbb{E}[(X_{t_{i+1}}-X_{t_i})^4|\mathscr{F}_{t_i}]. \tag{A.11}$$

In the other hand, we have

$$(Z_{i+1}-Z_i-\Delta\hat{\mu}(X_{t_i}(\omega)))^2 = (Z_{i+1}-Z_i)^2 - 2\Delta\hat{\mu}(X_{t_i}(\omega))(Z_{i+1}-Z_i) + (\Delta\hat{\mu}(X_{t_i}(\omega)))^2. \tag{A.12}$$

Using the result (1) and (A.10), we get almost surely that

$$\begin{aligned}\mathbb{E}[\mathbb{E}'[\frac{1}{\Delta}(Z_{i+1}-Z_i-\Delta\hat{\mu}(X_{t_i}(\omega)))^2]|\mathscr{F}_{t_i}] =& \mathbb{E}[\mathbb{E}'[\frac{1}{\Delta}(Z_{i+1}-Z_i)^2|\mathscr{F}_{t_i}]+\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2 \\ &-2(O_p\left(\frac{\Delta}{h}\right))\frac{1}{\Delta}\underbrace{\mathbb{E}[\mathbb{E}'[(Z_{i+1}-Z_i)]|\mathscr{F}_{t_i}]}_{=0} \\ &=\mathbb{E}[\mathbb{E}'[\frac{1}{\Delta}(Z_{i+1}-Z_i)^2|\mathscr{F}_{t_i}]+\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2 \\ &=\frac{1}{\Delta}\mathbb{E}[(X_{t_{i+1}}-X_{t_i})^2|\mathscr{F}_{t_i}]+\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2 \\ &-\frac{1}{\Delta}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)^2|\mathscr{F}_{t_i}\right]. \end{aligned} \tag{A.13}$$

Note that

$$\begin{aligned}\mathbb{E}[\frac{1}{\Delta}(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2|\mathscr{F}_{t_i}] =& \frac{1}{\Delta}\mathbb{E}[(X_{t_{i+1}}-X_{t_i})^2|\mathscr{F}_{t_i}]+\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2 \\ &-2(O_p\left(\frac{\Delta}{h}\right))\frac{1}{\Delta}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)|\mathscr{F}_{t_i}\right]. \end{aligned} \tag{A.14}$$

Combining the results (A.14) and (A.13) to obtain almost surely

$$\begin{aligned}\mathbb{E}[\mathbb{E}'[\frac{1}{\Delta}(Z_{i+1}-Z_i-\Delta\hat{\mu}(X_{t_i}(\omega)))^2]|\mathscr{F}_{t_i}] =& \mathbb{E}[\frac{1}{\Delta}(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2|\mathscr{F}_{t_i}] \\ &+2(O_p\left(\frac{\Delta}{h}\right))\frac{1}{\Delta}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)|\mathscr{F}_{t_i}\right]. \end{aligned} \tag{A.15}$$

Define

$$U^n(x)=\frac{\bar{T}}{nh}\sum_{i=0}^{n-1}K\left(\frac{X_{t_i}-x}{h}\right)\bar{\sigma}^2(t_i), \tag{A.16}$$

where

$$\bar{\sigma}^2(t_i)=\frac{1}{\Delta}(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2 \tag{A.17}$$

is computed at the trajectory of X_t. Then, we rewrite $S^n_{new}(x) = U^n(x)/L^n_{\bar{T}}(x)$. Note that $||X||^2 = \mathbb{E}[X^2]$ represents the $\mathscr{L}^2(\Omega)$ norm of X. Now, by using the identity (A.15), equation (A.13) and the properties of $\mathscr{L}^2(\Omega)-$norm to obtain

$$\begin{aligned}
||U^n(x) - \sigma^2(x)L^n_{\bar{T}}(x)|| =& \left|\left| \frac{1}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)\mathbb{E}'[(Z_{i+1}-Z_i-\Delta\hat{\mu}(X_{t_i}(\omega)))^2] - \sigma^2(x)L^n_{\bar{T}}(x) \right.\right. \\
&\left.\left. -2\frac{\bar{T}}{nh}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)(O_p\left(\frac{\Delta}{h}\right))\frac{1}{\Delta}\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)\right|\right| \\
\leq& \left|\left| \frac{1}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)\mathbb{E}'[(Z_{i+1}-Z_i-\Delta\hat{\mu}(X_{t_i}(\omega)))^2] - \sigma^2(x)L^n_{\bar{T}}(x)\right|\right| \\
&+ \left|\left| 2\frac{\bar{T}}{nh}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)(O_p\left(\frac{\Delta}{h}\right))\frac{1}{\Delta}\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)\right|\right| \\
\leq& \left|\left| \frac{1}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)(X_{t_{i+1}}-X_{t_i})^2 - \sigma^2(x)L^n_{\bar{T}}(x)\right|\right| \\
&+ \left|\left| 2\frac{\bar{T}}{nh}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)(O_p\left(\frac{\Delta}{h}\right))\frac{1}{\Delta}\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)\right|\right| \\
&+ \left|\left| \frac{\Delta}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2\right|\right| \\
&+ \left|\left| \frac{\Delta}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)\frac{1}{\Delta}\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)^2\right|\right|. \qquad \text{(A.18)}
\end{aligned}$$

Using the fact that $\mu(\cdot)$ is bounded from assumption 3.(iii), the inequality (A.37) yields

$$\begin{aligned}
||U^n(x) - \sigma^2(x)L^n_{\bar{T}}(x)| \leq& \left|\left| \frac{1}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)(X_{t_{i+1}}-X_{t_i})^2 - \sigma^2(x)L^n_{\bar{T}}(x)\right|\right| \\
&+ M\left|\left| 2\frac{\bar{T}}{nh}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)O_p\left(\frac{\Delta}{h}\right)\right|\right| \\
&+ \left|\left| \frac{\Delta}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2\right|\right| + M^2\Delta\left|\left| \frac{\Delta}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{t_i}-x}{h}\right)\right|\right|. \qquad \text{(A.19)}
\end{aligned}$$

All three terms at right-side of the inequality (A.39) converge to zero as n increases to infinity. The first term converges to zero because of the Proposition (3); and the second one converges to zero since $O_p\left(\frac{\Delta}{h}\right) \xrightarrow{p} 0$ as $n \to \infty$ and also because of the Proposition (2) or Lemma (2.25). The last term converges to zero because of the Proposition (2), $\Delta \to 0$ and also fact that $n^{1/2}h \to \infty$, by Slutsky's Theorem $\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2 \xrightarrow{p} 0$.

Thus, we conclude $U^n(x)/L^n(x)$ converges to $\sigma^2(x)$ in $\mathscr{L}^2$ sense and hence the proof of the Proposition (4) is completed.

Next, we recall the following result from Florens-Zmirou (1993) and extended by Jiang and Knight (1997).

Lemma 2. *Suppose* $\bar{T} = 1$. *If* $nh^3 \to 0$ *as* $n \to \infty$, *then*

$$\sum_{i=0}^{[nt]-1} \mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\left[(X_{t_{i+1}}-X_{t_i})^2-\sigma^2(x)/n\right]\right)^2 |\mathscr{F}_{t_i}\right] \to \sigma^4(x)L_t(x)$$

where the above convergence is in the probability.

Proof. We refer reader to the proof of the Lemma $(2b)$ in Florens-Zmirou (1993). □

Lemma 3. *Suppose* $\bar{T} = 1$. *Let* $g : \mathbb{R} \to \mathbb{R}$ *be a continuously differentiable bounded function, with bounded first derivative. Suppose that* $nh^3 \to 0$ *as* $n \to \infty$. *Consider*

$$G_t(x) = \frac{1}{\sqrt{nh}} \sum_{i=0}^{[nt]-1} \mathbb{E}\left[K\left(\frac{X_{t_i}-x}{h}\right)^2 |g(X_{t_i})-g(X_x)|\right].$$

Then, as $n \to \infty$, $G_t(x) \to 0$ *in the* $\mathscr{L}^1$ *sense and thus in probability.*

Remark 9. *Lemma 3 is similar to the Lemma* 3 *in Renò (2008). In our case we use the square Kernel function instead a kernel function.*

Proof. The proof of Lemma 3 is similar to one of Lemma 3 in Renò (2008) and therefore we omit the detail. □

Now, let us propose the following Lemma.

Lemma 4. *Suppose* $\bar{T} = 1$. *If* $nh^3 \to 0$ *as* $n \to \infty$, *then*

$$\sum_{i=0}^{[nt]-1} \mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\left[(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2-\sigma^2(x)/n\right]\right)^2 |\mathscr{F}_{t_i}\right] \to \sigma^4(x)L_t(x)$$

where the convergence is in the probability.

Proof. Using the identity relation (A.13) and almost surely we get

$$
\begin{aligned}
&\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\left[(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2-\sigma^2(x)/n\right]\right)^2\Big|\mathscr{F}_{t_i}\right]\\
&=\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\mathbb{E}'[(Z_{i+1}-Z_i)^2]+\left(O_p\Big(\frac{\Delta}{h}\Big)\right)^2-\sigma^2(x)/n\right]\right)^2\Big|\mathscr{F}_{t_i}\right]\\
&=\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\left[\mathbb{E}'[(Z_{i+1}-Z_i)^2]-\sigma^2(x)/n\right]\right)^2\Big|\mathscr{F}_{t_i}\right]\\
&+2\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)^2\left[\mathbb{E}'[(Z_{i+1}-Z_i)^2]-\sigma^2(x)\right]\right)\right.\\
&\left.\times\left(O_p\Big(\frac{\Delta}{h}\Big)\right)^2\Big|\mathscr{F}_{t_i}\right]+\mathbb{E}\left[\frac{n}{h}K\left(\frac{X_{t_i}-x}{h}\right)^2\left(O_p\Big(\frac{\Delta}{h}\Big)\right)^4\Big|\mathscr{F}_{t_i}\right]\\
&=\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\left[(X_{t_{i+1}}-X_{t_i})^2-\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)^2-\sigma^2(x)/n\right]\right)^2\Big|\mathscr{F}_{t_i}\right] \qquad \text{(A.20)}\\
&+2\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)^2\left[(X_{t_{i+1}}-X_{t_i})^2-\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)^2-\sigma^2(x)/n\right]\right)\right.\\
&\left.\times\left(O_p\Big(\frac{\Delta}{h}\Big)\right)^2\Big|\mathscr{F}_{t_i}\right] \qquad \text{(A.21)}\\
&+\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)^2\left(O_p\Big(\frac{\Delta}{h}\Big)\right)^4\Big|\mathscr{F}_{t_i}\right] \qquad \text{(A.22)}\\
&=A_{11}+A_{12}+A_{13}; \qquad \text{(A.23)}
\end{aligned}
$$

We want to examine the term A_{12}. Observe that

$$(X_{t_{i+1}}-X_{t_i})^2=X_{t_{i+1}}^2-X_{t_i}^2-2X_{t_i}(\Delta X_{t_{i+1}}-X_{t_i}). \qquad \text{(A.24)}$$

Note that by Itô′s formula we have

$$dX_t^2=2X_t\mu(X_t)dt+2X_t\sigma(X_t)dB_t+\sigma(X_t)^2dt, \qquad \text{(A.25)}$$

and thus we obtain

$$X_{t_{i+1}}^2 - X_{t_i}^2 = 2\int_{t_i}^{t_{i+1}} X_s \mu(X_s) ds + 2\int_{t_i}^{t_{i+1}} X_s \sigma(X_s) dB_s + \int_{t_i}^{t_{i+1}} \sigma(X_s)^2 ds. \tag{A.26}$$

Using both (A.24) and (A.26), and a direct calculation leads to

$$\begin{aligned}
\Delta X_{t_i}^2 &= (X_{t_{i+1}} - X_{t_i})^2 \\
&= 2\int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) \mu(X_s) ds + 2\int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) \sigma(X_s) dB_s \\
&+ \int_{t_i}^{t_{i+1}} \sigma(X_s)^2 ds.
\end{aligned} \tag{A.27}$$

Substituting (A.27) into (A.21) and A_{12} yields

$$\begin{aligned}
A_{12} = &\left(O_p\left(\frac{\Delta}{h}\right)\right)^2 \mathbb{E}\left[\frac{2n}{h}\left(K\left(\frac{X_{t_i} - x}{h}\right)^2 \left[2\int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) \mu(X_s) ds\right.\right. \\
&\left.\left. + \int_{t_i}^{t_{i+1}} (\sigma(X_s)^2 - \sigma^2(x)) ds - \left(\int_{t_i}^{t_{i+1}} \mu(X_s) ds\right)^2\right]\right) \Big| \mathscr{F}_{t_i}\right] \\
&+ \left(O_p\left(\frac{\Delta}{h}\right)\right)^2 \mathbb{E}\left[\frac{2n}{h}\left(K\left(\frac{X_{t_i} - x}{h}\right)^2 \left[2\int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) \sigma(X_s) dB_s\right]\right) \Big| \mathscr{F}_{t_i}\right] \qquad \text{(A.28)} \\
= &\left(O_p\left(\frac{\Delta}{h}\right)\right)^2 \mathbb{E}\left[\frac{2n}{h}\left(K\left(\frac{X_{t_i} - x}{h}\right)^2 \left[2\int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) \mu(X_s) ds\right.\right. \\
&\left.\left. + \int_{t_i}^{t_{i+1}} (\sigma(X_s)^2 - \sigma^2(x)) ds - \left(\int_{t_i}^{t_{i+1}} \mu(X_s) ds\right)^2\right]\right) \Big| \mathscr{F}_{t_i}\right] \qquad \text{(A.29)}
\end{aligned}$$

The expression (A.29) converges to zeros since it martingale. Now, Let us examine the term (A.29). Suppose that

$$a_{11}^{\#} = K\left(\frac{X_{t_i} - x}{h}\right)^2 \left[2\int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) \mu(X_s) ds + \int_{t_i}^{t_{i+1}} (\sigma(X_s)^2 - \sigma^2(x)) ds - \left(\int_{t_i}^{t_{i+1}} \mu(X_s) ds\right)^2\right]$$

Set $\mathbb{E}_i = \mathbb{E}[\cdot|\mathscr{F}_{t_i}]$. Taking the conditional expected value of $a_{11}^{\#}$ a

$$\begin{aligned}\mathbb{E}_i[|a_{11}^{\#}|] =&\mathbb{E}_i\left[K\left(\frac{X_{t_i}-x}{h}\right)^2\left|2\int_{t_i}^{t_{i+1}}(X_s-X_{t_i})\mu(X_s)ds+\int_{t_i}^{t_{i+1}}(\sigma(X_s)^2-\sigma^2(x))ds-\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)^2\right|\right]\\ &\leq \mathbb{E}_i\left[K\left(\frac{X_{t_i}-x}{h}\right)^2\left[2\int_{t_i}^{t_{i+1}}|X_s-X_{t_i}||\mu(X_s)|ds+\int_{t_i}^{t_{i+1}}|\sigma(X_s)^2-\sigma^2(x)|ds\right.\right.\\ &\left.\left.+\left(\int_{t_i}^{t_{i+1}}|\mu(X_s)|ds\right)^2\right]\right].\end{aligned}$$

As in Bandi and Phillips (2003) or Florens-Zmirou (1993), by applying the Burkholder Davis-Gundy inequality (see Section 4, Exercise 4.13 in Revuz and Yor (1999)) to the following quantity

$$\mathbb{E}_i\left[\sup_{s\leq(i+1)\Delta_{n,\bar{T}}}|X_s-X_{t_i}|\right] \tag{A.30}$$

and also applying Cauchy-Schwarz inequality, we get

$$\begin{aligned}\mathbb{E}_i[|a_{11}^{\#}|] \leq& K\left(\frac{X_{t_i}-x}{h}\right)^2\left\{\left(\left[\mathbb{E}_i\left[\sup_{s\leq(i+1)\Delta_{n,\bar{T}}}2(X_s-X_{t_i})\right]^2\right)^{\frac{1}{2}}\left(\mathbb{E}_i\left[\int_{t_i}^{t_{i+1}}|\mu(X_s)|ds\right]^2\right)^{\frac{1}{2}}\right.\\ &+\mathbb{E}_i\left(\int_{t_i}^{t_{i+1}}|\sigma(X_s)^2-\sigma^2(X_{t_i})|ds\right)+\int_{t_i}^{t_{i+1}}|\sigma(X_{t_i})^2-\sigma^2(x)|ds+\left.\mathbb{E}_i\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)^2\right\}\\ \leq& K\left(\frac{X_{t_i}-x}{h}\right)^2\left\{\text{const1}.\frac{1}{\sqrt{n}}\frac{1}{n}+\text{cosnt}.\frac{1}{\sqrt{n}}\frac{1}{n}+\frac{1}{n}|\sigma(X_{t_i})^2-\sigma^2(x)|+\frac{M^2}{n^2}\right\}.\end{aligned} \tag{A.31}$$

where const1 and const are some suitable constants. The inequality (A.31) follows because of (A.30), and also the fact that $\mu(\cdot)$ bounded, $\sigma(\cdot)$ is bounded with the bounded first derivative. Thus, we bound the A_{12} by

$$A_{12}\leq\left(O_p\left(\frac{\Delta}{h}\right)\right)^2\frac{2}{h}K\left(\frac{X_{t_i}-x}{h}\right)^2\left\{\text{const1}.\frac{1}{\sqrt{n}}+\text{cosnt}.\frac{1}{\sqrt{n}}+|\sigma(X_{t_i})^2-\sigma^2(x)|+\frac{M^2}{n}\right\}. \tag{A.32}$$

Using the result (A.32) and the boundedness of $\mu(\cdot)$ we get

$$\begin{aligned}
&\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\left[(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2-\sigma^2(x)/n\right]\right)^2\Big|\mathscr{F}_{t_i}\right]\\
\leq&\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\left[(X_{t_{i+1}}-X_{t_i})^2-(M\Delta)^2-\sigma^2(x)/n\right]\right)^2\Big|\mathscr{F}_{t_i}\right]\\
&+\left(O_p\left(\frac{\Delta}{h}\right)\right)^2\frac{2}{h}K\left(\frac{X_{t_i}-x}{h}\right)^2\left\{\text{const1.}\frac{1}{\sqrt{n}}+\text{cosnt.}\frac{1}{\sqrt{n}}+|\sigma(X_{t_i})^2-\sigma^2(x)|+\frac{M^2}{n}\right\}\\
&+\left(O_p\left(\frac{\Delta}{h}\right)\right)^4\mathbb{E}\left[\frac{n}{h}K\left(\frac{X_{t_i}-x}{h}\right)^2\Big|\mathscr{F}_{t_i}\right].
\end{aligned}$$

Now, subtracting the both side of the above inequality by $\sigma^4(x)L_t(x)$ and take the expected value yields

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=0}^{n-1}\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\left[(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2-\sigma^2(x)/n\right]\right)^2\Big|\mathscr{F}_{t_i}\right]-\sigma^4(x)L_t(x)\right]\\
\leq&\mathbb{E}\left[\sum_{i=0}^{n-1}\mathbb{E}\left[\frac{n}{h}\left(K\left(\frac{X_{t_i}-x}{h}\right)\left[(X_{t_{i+1}}-X_{t_i})^2-(M\Delta)^2-\sigma^2(x)/n\right]\right)^2\Big|\mathscr{F}_{t_i}\right]-\sigma^4(x)L_t(x)\right]\\
&+\mathbb{E}\left[\sum_{i=0}^{n-1}\frac{2\Delta}{h}K\left(\frac{X_{t_i}-x}{h}\right)^2\right]\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2\left\{\text{const1.}\frac{1}{\sqrt{n}}+\text{cosnt.}\frac{1}{\sqrt{n}}+\frac{M^2}{n}\right\}\\
&+\left(O_p\left(\left(\frac{\Delta^3}{h^5}\right)^{\frac{1}{4}}\right)\right)^2\mathbb{E}\left[\sum_{i=0}^{n-1}\frac{2}{\sqrt{nh}}K\left(\frac{X_{t_i}-x}{h}\right)^2|\sigma(X_{t_i})^2-\sigma^2(x)|\right]\\
&+\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^4\mathbb{E}\left[\sum_{i=0}^{n-1}\frac{\Delta}{h}K\left(\frac{X_{t_i}-x}{h}\right)^2\right].
\end{aligned}$$

All for terms at the right-side of the above inequality converge to zero in probability sense as n increases to infinity. The first term converges to zero because of the Lemma 2 and also fact that $\Delta\to 0$ as $n\to\infty$. The second and last terms converge to zeros because of Proposition 2, and Assumption 5.(ii). The third term converges to zero because of Lemma 3 and Assumption 5.(ii). Thus, we get the result. □

Let us now prove the Proposition 5.

A.3 Proof of Proposition 5

We note that :

$$\sqrt{nhL_t^n(x)}\left(\frac{S_{new}^n(x)}{\sigma^2(x)}-1\right)=\frac{\sum_{i=0}^{n-1}H_i}{\sigma^2(x)\sqrt{L_t^n(x)}}$$

where

$$H_i=\frac{\bar{T}}{\sqrt{nh}}K\left(\frac{X_{t_i}-x}{h}\right)\eta_{i+1} \tag{A.33}$$

$$\eta_{i+1}=\bar{\sigma}^2(X_{t_i})-\sigma^2(x)$$

$$\bar{\sigma}^2(X_{t_i})=\Delta^{-1}(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2. \tag{A.34}$$

Since $\bar{\sigma}^2(\cdot)$ is bounded, $H(\cdot,\cdot)$ is bounded and it is adapted to $\mathscr{F}_{t_i}$. To prove the Proposition 5 we need to verify the following (cf: Lemma 3 in Renò (2008)):

(i) $\sum_{i=0}^{n-1}\mathbb{E}[H_i|\mathscr{F}_{t_i}]\to 0$ in probability

(ii) $\sum_{i=0}^{n-1}\mathbb{E}[H_i^2|\mathscr{F}_{t_i}]\to\sigma^4(x)L_t(x)$ in probability

(iii) For every $\varepsilon>0$, $\sum_{i=0}^{n-1}\mathbb{E}[H_iI_{H_i^2>\varepsilon}|\mathscr{F}_{t_i}]\to 0$ in probability (conditional Lindeberg condition)

Let us start by the conditions

(i) As the above prove, we set $\mathbb{E}[\cdot|\mathscr{F}_{t_i}]=\mathbb{E}_i[\cdot]$ and $\sum_{i=0}^{n-1}=\sum$. Then, we have

$$\begin{aligned}\mathbb{E}_i[H_i]&=\mathbb{E}_i\left[\frac{\bar{T}}{\sqrt{nh}}K\left(\frac{X_{t_i}-x}{h}\right)\left[\Delta^{-1}(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2-\sigma^2(x)\right]\right]\\&=\mathbb{E}_i\left[\sqrt{\frac{n}{h}}K\left(\frac{X_{t_i}-x}{h}\right)\left[(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2-\Delta\sigma^2(x)\right]\right]\end{aligned}$$

Using the content arguments for the proof of Lemma 4, we find

$$\begin{aligned}|\mathbb{E}[\sum\mathbb{E}_i[H_i]]|\leq&\sqrt{h}\left\{\text{const1}+\text{cosnt}+\frac{M^2}{\sqrt{n}}\right\}\mathbb{E}[L_t^{(n)}(x)]\\&+\mathbb{E}\left[\sum\frac{1}{\sqrt{nh}}K\left(\frac{X_{t_i}-x}{h}\right)|\sigma(X_{t_i})^2-\sigma^2(x)|\right].\end{aligned}$$

All terms at the right-side of the inequality converge to zeros in probability. The first term converges to zero because $L_t^{(n)}(x)$ converges to $L_t(x)$ and h tends zero. The last term converges to zero because of Lemma 3 in Renò (2008).

(ii) we need to prove $\sum_{i=0}^{n-1}\mathbb{E}[H_i^2|\mathscr{F}_{t_i}] \xrightarrow{p} \sigma^4(x)L_t(x)$. Using Lemma 4, this is equivalent to show that

$$\Lambda_t = \sum \mathbb{E}_i\left[\frac{T^2}{nh}\left\{K\left(\frac{X_{t_i}-x}{h}\right)\left[\Delta^{-1}(X_{t_{i+1}}-X_{t_i}-\Delta\hat{\mu}(X_{t_i}))^2-\Delta^{-1}(X_{t_{i+1}}-X_{t_i})^2\right]\right\}^2\right] \xrightarrow{p} 0.$$

Now, by using the content arguments of the proof of Proposition 4 together with identity (A.13)

$$\begin{aligned}\Lambda_t &= \sum \mathbb{E}_i\left[\frac{T^2}{nh}\left\{K\left(\frac{X_{t_i}-x}{h}\right)\left[\Delta^{-1}\mathbb{E}'[Z_{i+1}-Z_i]-\Delta^{-1}(X_{t_{i+1}}-X_{t_i})^2\right.\right.\right.\\ &\left.\left.\left.-2(O_p\left(\frac{\Delta}{h}\right))\frac{1}{\Delta}\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)+\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2\right]\right\}^2\right]\\ &= \sum \mathbb{E}_i\left[\frac{T^2}{nh}\left\{K\left(\frac{X_{t_i}-x}{h}\right)\left[-\Delta^{-1}\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)^2\right.\right.\right.\\ &\left.\left.\left.-2(O_p\left(\frac{\Delta}{h}\right))\frac{1}{\Delta}\left(\int_{t_i}^{t_{i+1}}\mu(X_s)ds\right)+\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2\right]\right\}^2\right].\end{aligned}$$

Applying the triangle inequality property of $||\cdot||_{\mathscr{L}^2}$ and because of Assumption 3.(iii), we get

$$\Lambda_t \leq \sum \mathbb{E}_i\left[\frac{T^2}{nh}\left\{K\left(\frac{X_{t_i}-x}{h}\right)\Delta M^2\right\}^2\right] \tag{A.35}$$

$$+\sum \mathbb{E}_i\left[\frac{T^2}{nh}\left\{K\left(\frac{X_{t_i}-x}{h}\right)\left[-2(O_p\left(\frac{\Delta}{h}\right))M+\left(O_p\left(\frac{\sqrt{\Delta}}{h}\right)\right)^2\right]\right\}^2\right]. \tag{A.36}$$

All terms at the right hand side of the inequality converges to zero in probability. The first one converges to zero because of Proposition 2 and fact that Δ converges to zero. The second one converges to zero because of Proposition 2, Assumption 5.(ii) and fact that $\frac{\Delta}{nh}$ converges to zero. Hence we conclude the result.

(iii) Let us now prove the conditional Lindeberg condition. We have

$$\mathbb{E}[|\mathbb{E}_i[H_i^2 I_{|H_i|>\varepsilon}]|] = \sum_{|H_i|>\varepsilon} H_i^2. \tag{A.37}$$

Note that the sum of (A.37) is bounded by $\sigma(x)^4 L_t(x)$. To verify the conditional Lindeberg condition, first use fact that $|H_i| > \varepsilon$ and then we can write

$$K\left(\frac{X_{t_i}-x}{h}\right)|\eta_{i+1}| \geq \varepsilon\frac{\sqrt{nh}}{\bar{T}}, \tag{A.38}$$

where η_{i+1} is given in Eq.(A.33). Let us set

$$H_i' = K\left(\frac{X_{t_i}-x}{h}\right)\eta_{i+1}, \tag{A.39}$$

and we have

$$\mathbb{E}[|\mathbb{E}_i[H_i^2 I_{|H_i|>\varepsilon}]|] = \mathbb{E}[|\mathbb{E}_i[H_i^2 I_{|H_i'|>\varepsilon\frac{\sqrt{nh}}{\bar{T}}}]|]. \tag{A.40}$$

To simplify the notation, let Y_i denote the following expression $H_i^2 I_{|H_i'|>\varepsilon\frac{\sqrt{nh}}{\bar{T}}}$. Thus, we wish to prove that $\mathbb{E}[\mathbb{E}_i[Y_i]] \to 0$. Note that the equation (A.39) is bounded; thus as n tends to infinity we have $\varepsilon\sqrt{nh} \to \infty$ and Y_i vanishes in probability. We can also see that $|Y_i| \leq H_i^2$; and we know $\mathbb{E}[\sum|\mathbb{E}_i[H_i^2]|] < \infty$. Therefore, by invoking the Dominated Convergence Theorem we fund that $\sum\mathbb{E}[\mathbb{E}_i[Y_i]] \to 0$ and hence we conclude the sum of (A.37) converges to zero. Thus, we fulfil the assumption stated above.

Now, if one defines $Z_t^n(x)$ like

$$Z_t^n(x) = \sum H_i(x), \tag{A.41}$$

then one has that $Z_t^n(x)$ converges in law to the continuous martingale M_t with quadratic variation given by $[M,M]_t = \sigma^4(x)L_t(x)$.[4] We then set $M_t = B(\sigma^4(x)L_t(x))$ where $B(t)$ is a standard Brownian motion. Now, let us consider

$$\Theta_t^n(x) = \sum(W_{t_{i+1}} - W_{t_i}) \tag{A.42}$$

[4]Here we adapt the similar reasoning as in the proof of Theorem 2 in Renò (2008).

where W_t is a same standard Brownian motion in (2.6). We can clearly see that $\Theta_t^n(x)$ converges in law to the Brownian motion W_t. Furthermore, we have

$$\sum \mathbb{E}_i[H_i(W_{t_{i+1}} - W_{t_i})] = 0. \tag{A.43}$$

From equation (A.43), we find that $B(t)$ and W_t are orthogonal. Let us also write $B(t) = M_{\upsilon(t)}$ where $\upsilon(t) = \inf_s(\sigma^4(x)L_s(x))$. Now, by invoking the Knight's theorem (see Knight (1971)), we find that $B(t)$ and W_t are independent Brownian motion. Then $B(t)$ and $L_t(x)$ are independent, using the fact that the filtration generated by X_t is included in the one generated by W_t. We then have that $Z_t^n(x) \to \sigma^2(x)\sqrt{L_s(x)}\mathcal{N}(0,1)$, where $\mathcal{N}(0,1)$ is a standard Brownian motion independent to $L_t(x)$. Now, using fact that $L_t^{(n)}(x)$ converges in probability to $L_t(x)$, then we find that $\frac{Z_t^n(x)}{\sigma^2(x)\sqrt{L_s^n(x)}} \to \mathcal{N}(0,1)$. Thus, we conclude the proof.

Chapter 3

Realized Factor Quasi-Maximum-Likelihood Estimation of Large Covariance Matrices

3.1 Introduction

The covariance matrix of multiple assets plays a key role in many financial applications such as portfolio optimization, risk management and option pricing. The availability of high-frequency data has led to the development of several realized covariance (RC) estimators of the so-called integrated covariance (IC). The latter can be regarded as a measure of the covariance among asset prices when the latter are described by a continuous-time semimartingale, see Barndorff-Nielsen and Shephard (2004).

Unfortunately, unlike the low-frequency data which are homogeneously spaced, high-frequency (tick-by-tick) data are irregularly spaced. More precisely, the market data samples of an asset are randomly spaced in time and transactions of two or more assets occur asynchronously. In addition, high-frequency data are contaminated by microstructure noise (bid/ask spread). These factors make the estimation of covariance matrix using high-frequency data a difficult task. For instance, if we synchronise the data by previous-tick interpolation and apply the realized covariance estimator of Barndorff-Nielsen and Shephard (2004), this will result in a downward bias on the estimated covariance (Hayashi and Yoshida 2005). The problem of the stock correlations decreasing when the sampling frequency increases is known as “Epps Effect" and was first reported by Epps (1979).

To tackle these problems, several approaches have been proposed in the literature (see, Bollerslev and Zhang 2003, Renò 2003, Griffin and Oomen 2011, Barndorff-Nielsen et al. 2011, Mancino and Sanfelici 2011, Zhang 2011, Hautsch

et al. 2012, Corsi et al. 2015 and Shephard and Xiu 2017). The aforementioned covariance estimators can be grouped into two categories, *non-parametric* and *parametric estimators*. Most non-parametric methods for realized cavariance matrix, e.g. the multivariate realized kernel of Barndorff-Nielsen et al. (2011), require the use of linear interpolation, previous-tick interpolation, or other interpolation schemes to create a synchronized pair of asset prices in case data are irregularly spaced. Therefore, they can not be applied directly to intraday data. This effect can lead to ignoring much of the data and can significantly jeopardize the estimation of the covariance matrix. In contrast to these non-parametric estimators, the parametric estimators for realized covariance, like the Quasi-Maximum Likelihood Estimator (QMLE) of Corsi et al. (2015) and Shephard and Xiu (2017) do not require the use of synchronisation schemes. The QMLE is based on a state-space representation of high-frequency data. This allows to treat asynchronicity as a standard missing value problem in state-space models. As such, the QMLE uses the all the available data when estimating the covariance. Furthermore, as the state-space representation is linear and Gaussian, estimation is performed by a standard quasi-maximum likelihood method.

However, when the dimension of the number of asset (d) increases to a few hundreds, it becomes difficult to estimate the integrated covariance matrix using QMLE. The main problem is that the number of parameters increase as $\mathcal{O}(d^2)$, and, moreover, as d goes to infinity, the Kalman filter becomes computationally infeasible. This problem is known as the curse of dimensionality and might pose great challenges when estimating covariance matrices of high-dimensions.

In order to tackle curse of dimensionality, one can rely on factor models. The underlying idea of factor models is that the comovements of financial markets can be captured by few factors. Prominent examples are given by the capital asset pricing model (CAPM); see, for example, Markowitz (1952); Sharpe (1964); Black (1972), and Lintner (1975), and the three-factor model of Fama and French (1993). The aforementioned methods assume that the factors are observable. When the factors are latent, state-space and principal component analysis (PCA) methods are the main tools to estimate the dynamic factor models. There is a longstanding literature on this topic. Some examples are: Geweke (1977), Sargent and Sims (1977), Stock and Watson (2002), Giannone et al. (2005), Giannone et al. (2008), Doz et al. (2011), Doz et al. (2012) and Bańbura and Modugno (2014). The above-mentioned estimators have been developed for low frequency data. Recently, Ait-Sahalia and Xiu (2017), Ait-Sahalia and Xiu (2019), and Pelger (2019a) developed principal component analysis (PCA)-based inferential tools for functionals of semimartingales. These estimators are based on non-parametric estimators of the integrated covariance. However, since the non-parametric estimator is not robust to asynchrounous trading, we must use the previous-tick synchronization or related interpolation schemes to reconstruct the log-prices. As indicated above, this procedure leads to a downward bias of covariance towards zero and, therefore, could also affect PCA-based methods, since the latter rely on the estimation of the realized covariance matrix. Contrary to non-parametric methods, parametric state-space methods, like in Bańbura and Modugno (2014) and Creal et al. (2014) have a clear advantage in dealing with

asynchronous trading, since missing values can be integrated out when computing the log-likelihood. However, Bańbura and Modugno (2014) and other related works on factor models applied to low-frequency data, model the dynamic of *log-returns* rather than *log-prices*. It should be noted that by modeling the dynamic of log-returns, we might face a great difficult when estimating the covariance matrix from high-frequency data, because in the presence of missing values in log-prices, models of log-returns are subject to significant data reduction. Suppose that an asset is traded at time t, but not at time $t-1$, a situation that is very common at the intraday level. A parametric model for log-returns would then treat the return at t as a missing value and completely ignore the log-price information at t. In contrast, if we model the dynamic of log-prices instead of log-returns, the problem of data reduction can be overcome when estimating the high-dimensional covariance matrix.

In the literature no effort has been devoted to exploiting the factor structure of *intraday data* with the parametric methods. As already stated, the parametric state-space methods have a clear advantage in dealing with asynchronous trading. Asynchronous data can be treated as missing observations, which are integrated out when computing the log-likelihood. In this work, we aim to fill the gap in the literature.

We tackle the problem of dimensionality reduction in the estimation of integrated covariances in the presence of market microstructure effects and non-synchronous trading. We propose to model intraday prices through a dynamic factor model where idiosyncratic errors incorporate microstructure effects. Estimation is performed by assuming a linear-Gaussian state-space representation and employing an Expectation-Maximization algorithm. The model provides a fully parametric setup where one can treat asynchronicity as a missing data problem, in a similar fashion to Corsi et al. (2015) and Shephard and Xiu (2017). The advantages are that covariances can be estimated using the entire informational content of the data and the problem of model selection, namely how to choose the correct number of factors, is easily addressed through a likelihood-based criterion. The advantage, compared to QMLE, is that the proposed dynamic factor models, named "realized factor QMLE" is not affected by the curse of dimensionality.

We perform an extensive Monte Carlo simulation study to investigate the statistical properties of the realized factor QMLE. In the representation of the proposed dynamic factor models, we assume that the instantaneous covariance of the common factors is constant over time. Our Monte Carlo simulations show the accuracy of the estimator in recovering the true covariance matrix in the presence of stochastic volatility in the factors and asynchronous trading. We find that even in the presence of stochastic volatility, the estimator remains asymptotically unbiased. The obtained result is similar to that of Shephard and Xiu (2017) [1]. Empirically, we analyze the 100 most liquid assets traded in the NYSE, from January 3*rd*, 2006 to December 31*st*, 2014. In particularly, we study the factor structure of intraday prices and perform an horse race

[1] Shephard and Xiu (2017) provided the asymptotic theory of the QMLE estimator of the integrated covariance matrix of a Brownian semimartingale observed with noise

based on out-of-sample minimum variance portfolios to compare our estimator to standard integrated covariance estimators. We find evidence on the existence of a factor structure in high-frequency data. The results show that the dynamic factor specification does not lead to a significant increase of portfolio variances compared to the full model with no factor structure (QMLE). Moreover, the realized factor QMLE with number of common factor set to $20, 15, 10$ and 5 (QMLE $_{20}$, QMLE$_{20}$, QMLE$_{15}$, QMLE$_{10}$, QMLE$_5$) and QMLE$_{\hat{r}_{bic}}$ [2] outperform the two benchmark methods (PCA base-method of Ait-Sahalia and Xiu 2017 and the Multivariate Realized Kernel (MRK) of Barndorff-Nielsen et al. 2011). In forecasting the 1% and 5% VaR, the realized factors QMLE are statically indistinguishable from the QMLE and they outperform the PCA and MRK. By comparing the average estimation time for realized factor QMLE and QMLE, we observe that the realized factor QMLE perform much faster than the QMLE. In particular, the QMLE$_5$ is 12 times faster than the QMLE. Not surprisingly, if the number of factors r is small, for example 5 or 10, the calculation time required for the realized factor QMLE to estimate the covariance matrix is very small compared to that required by the QMLE. If r is large, say 20, the gains are less spectacular but nevertheless substantial.

The remainder of Chapter is organized as follows: In Section 3.2, we introduce the model setup, and then highlight the advantages of the proposed dynamic factor model type. In Section 3.3, we derive the realized factor QMLE, and Section 3.4 presents the Monte Carlo results. In Section 3.5, we describe the data and present the empirical results. Finally, Section 3.6 concludes the Chapter.

[2]The QMLE$_{\hat{r}_{bic}}$ is the realized factor QMLE with the number of common factors estimated on entire sample of 2265 trading days using Bayesian Information criterion (BIC).

3.2 The Model

Let $[0,T]$ denote a time interval. To fix ideas, $[0,T]$ can be thought of as representing the trading day. We consider n equispaced observations Y_{t_i}, $i = 1,\dots,n$ of a $d \times 1$ vector of intraday log-prices. We assume that the dynamics of Y_{t_i} can be modeled as:

$$\begin{cases} Y_{t_i} = HG_{t_i} + \eta_{t_i} & \eta_{t_i} \sim \mathcal{N}(0,R) \\ G_{t_i} = G_{t_{i-1}} + \xi_{t_i} & \xi_{t_i} \sim \mathcal{N}(0,\Lambda), \end{cases} \tag{3.1}$$

where G_{t_i} is an $r \times 1$ vector of common factors and H is a factor loading matrix of dimension $d \times r$. The two idiosyncratic errors η_{t_i} and ξ_{t_i} are assumed to be i.i.d. zero-mean normal distributed with covariance matrices R and Λ, respectively. The covariance matrix R is assumed to be diagonal. The common factors G_{t_i} represent a reduced number of "common trends" driving the components of the observed log-price Y_{t_i}. The full QMLE method of Corsi et al. (2015) and Shephard and Xiu (2017) is obtained for $r = d$.

As it is well-known, when $d \geq r$, i.e. the number of assets d is much larger than that of factors r, the number of parameters to be estimated is $\mathcal{O}(d)$ rather than $\mathcal{O}(d^2)$. Therefore, compared to the QMLE method of Corsi et al. (2015) and Shephard and Xiu (2017), the dynamic factor model in Eq. (3.1) is not affected by the curse of dimensionality. Note also that, due to asynchronous trading, only the components of Y_{t_i} corresponding to traded assets are observed at time t_i. As a result, there are a lot of missing values in the time-series of observed prices.

The use of standard methods to extract factors in low-frequency data can be extremely difficult in a high-frequency setting because of missing values. The main difficulty lies in the imputation of log-returns from the missing log-prices. There are two possible methods to deal with this problem. The first is to reconstruct log-prices using previous-tick synchronization or related interpolation schemes. A drawback of this method is that it leads to a large number of "artificial" zero returns, which are known to jeopardize the inference of realized covariances (see e.g. Epps 1979, Hayashi and Yoshida 2005, Buccheri et al. 2018). This effect might be crucial in PCA-based methods (see e.g. Stock and Watson 2002, Bai and Ng 2002, Fan et al. 2013, Ait-Sahalia and Xiu 2017, Pelger 2019a,b) since the latter rely on the estimation of the realized covariance matrix. A second method is the use of state-space models, like in Bańbura and Modugno 2014, Creal et al. 2014, among others. Missing values can easily be handled in state-space models. However, the above-mentioned works are based on a dynamic factor specification for *log-returns*. In the presence of missing values in *log-prices*, models for log-returns are subject to large data reduction. For instance, assume one asset is traded at time t, but is not traded at time $t-1$, a circumstance that is very common at the intraday level. A parametric model for log-returns would then treat the return at time t as a missing value and would completely neglect the information related to the log-price at time t. In Eq.

(3.2) we propose a dynamic factor specification for log-prices rather than for log-returns. This choice avoids the problem of data reduction and allows to estimate common factors using all the available log-prices.

3.3 Estimation

3.3.1 Realized factor Quasi-Maximum Likelihood Estimator

Without loss of generality, we set $t_{i+1} - t_i = 1$ and re-write Eq.(3.1) as:

$$\begin{cases} Y_t = HG_t + \eta_t & \eta_t \sim \mathcal{N}(0,R) \\ G_t = G_{t-1} + \xi_t & \xi_t \sim \mathcal{N}(0,\Lambda). \end{cases} \tag{3.2}$$

Model (3.2) is a linear Gaussian state-space representation whose likelihood function can be written in closed form using the Kalman filter. To estimate the parameters we can rely on two approaches. First, we can maximize the likelihood function numerically by a quasi-Newton method. This approach has the disadvantage of being computationally expensive, as it requires inverting the Hessian matrix of the parameters, which has large dimension in multivariate models. A second approach is to apply the EM algorithm of Dempster et al. (1977). The EM algorithm is a general method for finding the maximum-likelihood estimates of model parameters when data is incomplete. The method iteratively alternates between two steps known as the expectation step (E-step) and the maximization step (M-step), respectively. The EM algorithm is particularly advantageous for multivariate models because it does not require the inversion of the Hessian matrix. We thus adopt it to estimate the parameters of the dynamic factor model in Eq. (3.2). For more details on the EM algorithm, we refer the reader to Dempster et al. (1977) and Shumway and Stoffer (1982).

Let us first assume that the data do not contain missing observations. In Section (3.3.3) we will show how to handle the case of missing observations. Denote by $\Omega = \{H, \Lambda, R\}$ the set of parameters to be estimated. Let $(\mathscr{Y}_n, \mathscr{G}_n)$ denote the set of complete data, where $\mathscr{Y}_n = (Y_1, \ldots, Y_n)$ is the set of the observed prices and $\mathscr{G}_n = (G_1, \ldots, G_n)$ is the set of latent common factors. The joint or complete log-likelihood of the data $(\mathscr{Y}_n, \mathscr{G}_n)$, denoted by $L(\mathscr{Y}_n, \mathscr{G}_n; \Omega)$, is given by:

$$\begin{aligned} -2\log L(\mathscr{Y}_n, \mathscr{G}_n; \Omega) =& K + \sum_{t=1}^{n} [\log|\Lambda| + (G_t - G_{t-1})^\top \Lambda^{-1} (G_t - G_{t-1})] \\ &+ \sum_{t=1}^{n} [\log|R| + (Y_t - HG_t)^\top R^{-1} (Y_t - HG_t)], \end{aligned} \tag{3.3}$$

where K is a constant. We can re-write Eq. (3.3) as:

$$\begin{aligned}-2\log L(\mathscr{Y}_n,\mathscr{G}_n;\Omega) =& K + n\log|\Lambda| + \mathrm{Tr}\{\Lambda^{-1}\sum_{t=1}^{n}(G_t - G_{t-1})(G_t - G_{t-1})^\top\} \\ &+ n\log|R| + \mathrm{Tr}\{R^{-1}\sum_{t=1}^{n}(Y_t - HG_t)(Y_t - HG_t)^\top\}. \end{aligned} \tag{3.4}$$

Let us introduce the following conditional means and conditional covariances, which are an output of the Kalman filter and smoothing recursions in Appendix B.1:

$$G_t^s = \mathrm{E}[G_t|\mathscr{Y}_s] \tag{3.5}$$

$$V_t^s = \mathrm{Cov}[G_t|\mathscr{Y}_s] \tag{3.6}$$

$$V_{t,t-1}^s = \mathrm{Cov}[G_t, G_{t-1}|\mathscr{Y}_s] \tag{3.7}$$

where $s \leq n$. Typically, G_t^{t-1} is known as *predictive filter*, G_t^t is known as *update filter* and, finally, G_t^n is known as *smoother*. The same nomenclature holds for the case of conditional covariances V_t^s and $V_{t,t-1}^s$.

The $(j+1)$-th iteration of the EM algorithm for the dynamic factor model (3.2) is performed as follows: in the E-step, we calculate the conditional expectation of the complete log-likelihood function (3.4) assuming the parameters $\hat{\Omega}^j$ estimated at step j:

$$F(\Omega|\mathscr{Y}_n;\Omega^j) \equiv E(-2\log L(\Omega)|\mathscr{Y}_n;\Omega^j).$$

We have:

$$\begin{aligned}
F(\Omega|\mathscr{Y}_n;\Omega^j) &= K + n\log|\Lambda| + n\log|R| \\
&+ \mathrm{Tr}\{\Lambda^{-1}\sum_{t=1}^{n}\Big(V_t^n + G_t^n(G_t^n)^\top - V_{t,t-1}^n - G_t^n(G_{t-1}^n)^\top - V_{t-1,t}^n \\
&- (G_{t-1}^n)^\top G_t^n + V_{t-1,t-1}^n + G_{t-1}^n(G_{t-1}^n)^\top\Big)\} \\
&+ \mathrm{Tr}\{R^{-1}\sum_{t=1}^{n}\Big((Y_t - HG_t^n)^\top(Y_t - HG_t^n) + HV_t^nH^\top\Big)\} \\
&= K + n\log|\Lambda^{-1}| + n\log|R^{-1}| \\
&+ \mathrm{Tr}\{\Lambda^{-1}\Big(S_{11} - S_{10} - S_{10}^\top + S_{00}\Big)\} \\
&+ \mathrm{Tr}\{R^{-1}\sum_{t=1}^{n}\Big((Y_t - HG_t^n)(Y_t - HG_t^n)^\top + HV_t^nH^\top\Big)\}
\end{aligned} \tag{3.8}$$

where

$$\begin{cases}
S_{11} = \sum_{t=1}^{n}(V_t^n + G_t^n(G_t^n)^\top) \\
S_{10} = \sum_{t=1}^{n}(V_{t,t-1}^n + G_t^n(G_{t-1}^n)^\top) \\
S_{00} = \sum_{t=1}^{n}(V_{t-1,t-1}^n + G_{t-1}^n(G_{t-1}^n)^\top).
\end{cases} \tag{3.9}$$

The terms V_t^n, G_t^n and $V_{t-1,t-1}^n$ are computed under the parameters R^j, Λ^j and H^j through the Kalman filter and smoother recursions. In the M-step, we compute the new parameters of R^{j+1}, H^{j+1} and Λ^{j+1} by maximizing Eq. (3.8) w.r.t. Ω:

$$\Omega^{j+1} = \arg\max_{\Omega} F(\Omega|\mathscr{Y}_n;\Omega^j). \tag{3.10}$$

Eq. (3.10) is equivalent to differentiating Eq.(3.8) w.r.t. R, Λ and H as:

$$\nabla_H F(\Omega|\mathscr{Y}_n;\Omega^j)\Big|_{H=H^{j+1}} = 0, \qquad \nabla_R F(\Omega|\mathscr{Y}_n;\Omega^j)\Big|_{R=R^{j+1}} = 0,$$
$$\nabla_Q F(\Omega|\mathscr{Y}_n;\Omega^j)\Big|_{Q=Q^{j+1}} = 0.$$

We start by solving the equations $\nabla_H F(\Omega|\mathscr{Y}_n;\Omega^j)\Big|_{H=H^{j+1}} = 0$ and $\nabla_Q F(\Omega|\mathscr{Y}_n;\Omega^j)\Big|_{Q=Q^{j+1}} = 0$. The following proposition is proven in Appendix B.2:

Proposition 6. *The solutions of the equations*

$$\nabla_H F(\Omega|\mathscr{Y}_n;\Omega^j)\Big|_{H=H^{j+1}}=0, \quad \textit{and} \quad \nabla_Q F(\Omega|\mathscr{Y}_n;\Omega^j)\Big|_{Q=Q^{j+1}}=0$$

are given by :

$$H^{j+1}=\sum_{t=1}^{n} Y_t (G_t^n)^\top \Big(\sum_{t=1}^{n}(V_t^n+G_t^n(G_t^n)^\top)\Big)^{-1} \tag{3.11}$$

and

$$\Lambda_r^{j+1}=\frac{1}{n}\Big(S_{11}-S_{10}-(S_{10})^\top+S_{00}\Big). \tag{3.12}$$

We then use the estimated value of H to solve $\nabla_R F(\Omega|\mathscr{Y}_n;\Omega^j)\Big|_{R=R^{j+1}}=0$. The proof of the following proposition is given in Appendix B.2:

Proposition 7. *The solution of the equations* $\nabla_R F(\Omega|\mathscr{Y}_n;\Omega^j)\Big|_{R=R^{j+1}}=0$, *is:*

$$R^{j+1}=diag\left(\sum_{t=1}^{n}\frac{(Y_t-H^{j+1}G_t^n)(Y_t-H^{j+1}G_t^n)^\top+H^{j+1}V_t^n(H^{j+1})^\top}{n}\right). \tag{3.13}$$

The two steps (E-steps and M-steps) are repeated using the updated parameters until the incomplete log-likelihood function (3.14) converges, i.e. it stabilizes at some fixed point. After an E- step and a subsequent M-step, the incomplete log-likelihood function must increase. To verify convergence of the EM algorithm, we follow the Wu et al. (1983) [3] approach and stopping the algorithm once the relative increase of the log-likelihood is lower than some small threshold. More precisely, let $L(\mathscr{Y}_n;\Omega^j)$ denote the log-likelihood function of step j:

$$-2\log L(\mathscr{Y}_n;\Omega^j)=n\log(\pi)+\sum_{t=1}^{n}\log|D^j|+\sum_{t=1}^{n}(Y_t-H^jG_t^{t-1})^\top D^{j-1}(Y_t-H^jG_t^{t-1}) \tag{3.14}$$

where $D^j=H^jV_t^{t-1}H^{j\top}+R^j$. The threshold is computed as:

$$\mu_j=\frac{|L(\mathscr{Y}_n,\mathscr{G}_n;\Omega^j)-L(\mathscr{Y}_n,\mathscr{G}_n;\Omega^{j-1})|}{|L(\mathscr{Y}_n,\mathscr{G}_n;\Omega^{j-1})|}. \tag{3.15}$$

We stop the algorithm when μ_j is lower than the threshold of 10^{-4}.

[3] Wu et al. (1983) study the conditions in which the EM algorithm converges to a local maximum of the incomplete log-likelihood function.

3.3.2 Identification of Dynamic factor model

It is well-known that the dynamic factor model in Eq. (3.2) is not identified if we do not assume additional restrictions on the parameters. Indeed, suppose that B is an $r \times r$ invertible matrix. The following dynamic factor model

$$\begin{cases} Y_t = H^* G_t^* + \eta_t & \eta_t \sim \mathcal{N}(0, R) \\ G_t^* = G_{t-1}^* + \xi_t^* & \xi_t^* \sim \mathcal{N}(0, B\Lambda B) \end{cases} \tag{3.16}$$

where $H^* = HB^{-1}$ and $G_t^* = BG_t$, is observationally equivalent to the original dynamic factor model in Eq. (3.2). To guarantee identification, several restrictions are possible. As in Bańbura and Modugno (2014), we impose:

$$H_{\text{rest}} = \begin{bmatrix} I_r \\ H^*_{\text{unrest}} \end{bmatrix} \tag{3.17}$$

where H^*_{unrest} is a $(d-r) \times r$ unrestricted matrix. The unrestricted matrix H^*_{unrest} is derived in Section (3.3) (see Proposition 6), and is given by:

$$H^*_{\text{unrest}} = \sum_{t=1}^{n} Y_t^* (G_t^n)^\top \Big(\sum_{t=1}^{n} (V_t^n + G_t^n (G_t^n)^\top) \Big)^{-1} \tag{3.18}$$

where $Y_t^* = [Y_{r+1,t}, \cdots, Y_{d,t}]$. Koopman et al. (2008), Proietti (2008) and Bai and Wang (2015) use the same restriction. We note that the restriction in Eq. (3.17) is based on the theoretical results of Geweke and Singleton (1981).

On the other hand, the fact that our method is based on the maximum likelihood approach, we can modify the M-step and derive in closed form the updating equation for the restricted loading. In fact, Bork et al. (2009) and Bork (2009) show how to modify the M-step of Watson and Engle (1983) in order to impose the restriction on the factor loading in the form of:

$$\mathrm{D_H vec(H)} = \kappa_{\mathrm{H}}, \tag{3.19}$$

where $\mathrm{vec(H)}$ is a $dr \times 1$ vector gathering all stacked columns of the restricted factor loading H, $\mathrm{D_H}$ is a $r^2 \times dr$ matrix containing zeros and ones, and κ_{H} is $r^2 \times 1$ vector containing zeros and ones. The number of rows of $\mathrm{D_H}$ (and κ_{H}) is equal to the number of restricted elements in H. According to Bork et al. (2009) and Bork (2009), the updating equation for the

restricted factor loading H subject to a linear restriction in Eq.(3.19) is given by:

$$\text{vec}(H_{\text{rest}}^{j+1}) = \text{vec}(H_{\text{unrest}}^{(j+1)}) + (S_{11}^{-1} \otimes R^j)\text{D}_\text{H}^\top(\text{D}_\text{H}(S_{11}^{-1} \otimes R^j)\text{D}_\text{H}^\top)^{-1}(\kappa_\text{H} - \text{D}_\text{H}\text{vec}(\text{H}_{\text{unrest}}^{j+1})) \tag{3.20}$$

where $\otimes$ is a tensor product. The matrices H_{unrest} and S_{11} are given in Eq. (3.11) and (3.9), respectively. The restricted factor loading in Eq. (3.20) is exactly the same as in Eq. (3.17). Therefore, we can use either the restricted factor loading in Eq. (3.20) or (3.17). When the observations vector Y_t contains some missing values, the same formula (3.17) or (3.20) is used with the unrestricted factor loading (3.18) replaced by that obtained in the case of missing values (see Eq. 3.21) .

3.3.3 Modification for missing values

The algorithm for Kalman filtering and smoothing in Appendix B.1 can be modified to take into account missing values. In particular, Shumway and Stoffer (1982), Harvey (1990) and Durbin and Koopman (2012) show how to modify the for Kalman filtering and smoothing recursions in the case of missing values. Assume that A_t is a $d_t \times d$ matrix where d_t represents the number of observations available at time t. Suppose that at time t some of the observations in Y_t are missing. In this case, the actual observed values of Y_t are given by $Y_t^* = A_t Y_t$ and, therefore, the first equation of the dynamic factor (3.2) is replaced by the following:

$$Y_t^* = H_t^* G_t + \eta_t^* \quad \eta_t^* \sim \mathcal{N}(0, R_t^*),$$

where $H_t^* = A_t H, \eta_t^* = A_t \eta_t$ and $R_t^* = A_t R A_t^\top$. The matrix A_t acts as a selection matrix because it deletes the missing data so that only the available data are used in the calculations. If, at time t, the vector Y_t does not contain the missing values, the matrix A_t is equal to identity matrix. Following Durbin and Koopman (2012), the algorithm for Kalman filtering and smoothing in Appendix (B.1) and the prediction error decomposition form of the log-likelihood, Eq.(3.14) remain valid, provided that we replace Y_t, H_t and R_t with Y_t^*, H_t^* and R_t^*. Note that the dimensions of the observation vector and matrices in the dynamic factor model (3.2) vary at each point in time and depend on the number of available prices. When Y_t contains missing observations, the above update formulas of factor loading in Eq. (3.11) and covariance matrix in Eq. (3.13) can no longer be used. To derive the updating equation for the factor loading and covariance matrix in this case, we need to modify the the maximization step to take into account the missing observations. When we take the conditional expectation of Eq. (3.4), we are only interested on the third term since the first two terms depend only on G_t^n, V_t^n and $V_{t,t-1}^T$. In the following Proposition, we provide the update formulas of the factor loading matrix and covariance matrix in the case of missing observations:

Proposition 8. *Suppose that I_d is the $d \times d$ identity matrix. In the presence of missing observations, the update formulas of the factor loading and covariance matrices are given by:*

$$vec(H^{(j+1)}) = \Big(\sum_{t=1}^{n}(V_t^n + G_t^n(G_t^n)^\top) \otimes A_t\Big)^{-1} vec\Big(\sum_{t=1}^{n} A_t Y_t (G_t^n)^\top\Big) \tag{3.21}$$

and

$$R^{j+1} = diag\left(\frac{1}{n}\sum_{t=1}^{n}\left\{(A_tY_t - A_tH^{j+1}G_t^n)(A_tY_t - A_tH^{j+1}G_t^n)^\top + A_tH^{j+1}V_t^n(H^{j+1})^\top A_t + A_t^* R^j (A_t^*)^\top\right\}\right) \tag{3.22}$$

where $A_t^* = I_d - A_t$.

Proof. The proof of Proposition 8 is sketched in Appendix B.3, see also e.g. Bańbura and Modugno (2014). □

Note that if at time t, the vector Y_t does not contain missing values, the updated factor loading matrix in Eq. (3.21) and covariance matrix in Eq. (3.22) coincide with the factor loading (3.11) and covariance matrix (3.13) obtained in the case with no-missing values, since the matrix A_t is equal to identity matrix I_d.

3.3.4 AIC and BIC

In general, the number of common factors for the dynamic factor model in Eq.(3.2) is unknown. To determine the number of common factors for the dynamic factor model in Eq. (3.2), we fit a set of dynamic factor models with different number of factors to the sample data. For each fitted model, we calculate the log-likelihood function in Eq. (3.14) along with the number of free parameters (model complexity). Then, we select the number of common factors as the one providing the best balance between model fit and complexity (see Bulteel et al. 2013). According to Billah et al. (2005), the proper way of conducting the model selection is to use the information criteria. In this work, we consider the two well-known information criteria, namely the *Akaike Information Criterion (AIC; Akaike 1974b)* and *Bayesian Information Criterion* (BIC; Schwarz et al. 1978). Let fp_r denote the number of free parameters for the dynamic factor model in Eq. (3.2), calculated as follows:

$$fp_r = n(r+1) - \frac{r(r-1)}{2}. \tag{3.23}$$

The AIC criterion penalizes the minus two times the log likelihood (see Eq. 3.14) by the number of free parameters and is calculated as follows:

$$AIC_r = -2\log L(\hat{\Omega}, Y_n) + 2fp_r, \tag{3.24}$$

where $\hat{\Omega} = \{\hat{H}, \hat{R}, \hat{\Lambda}\}$ and $Y_n = \{Y_1 \cdots, Y_n\}$. The first term at the right hand side of Eq. (3.24) is given by Eq. (3.14). The dynamic factor model in Eq. (3.2) with the lowest AIC value is considered to be the best. Thus, we select the number of factors of the model with the lowest AIC value. The drawback of the AIC criterion is that it tends to favour the model with higher number of free parameters (see, i.e. Cavanaugh and Shumway 1997). To address this issue, Shumway (1988) recommended the use of the BIC criterion:

$$BIC_r = -2\log L(\hat{\Omega}, Y_n) + fp_r \log(n), \tag{3.25}$$

where n is a sample size. The only difference with the AIC criterion is that the "penalty" term is modified. Unlike AIC criterion, the BIC criterion takes into account the sample size. The advantage of using the BIC criterion is that the probability of selecting the correct model is greater when the sample size increases (see Ceulemans and Van Mechelen 2005).

3.3.5 Benchmark: Principal component analysis to high-frequency data

In this work we consider PCA-based methods as one of our benchmarks. Thus, in this section we briefly discuss how the high dimensional covariance estimator based on the PCA is constructed. Let $\Delta Y_i^n = Y_{i\Delta_n} - Y_{(i-1)\Delta_n}$ denote the observed log-returns at the sampling frequency Δ_n. The PCA is based on a principal component decomposition of the realized covariance matrix estimator, given below

$$\widehat{\Sigma} = \sum_{i=1}^{n} (\Delta^n Y_i)(\Delta^n Y_i)^\top. \tag{3.26}$$

Suppose that $\hat{\lambda}_1 > \hat{\lambda}_2 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_d$ are the sample eigenvalues of $\hat{\Sigma}$ and that $\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_d$ are the corresponding eigenvectors. Let $\hat{r}$ be the estimator of the common factors r to be introduced below. If $\hat{r}$ is known, we can write the spectral decomposition of $\hat{\Sigma}$ as follows:

$$\widehat{\Sigma} = \widehat{\Sigma}_{\hat{r}}^{S} + \widehat{\Gamma}_{\hat{r}}, \tag{3.27}$$

where $\widehat{\Gamma}_{\hat{r}} = \sum_{i=\hat{r}+1}^{n} \hat{\lambda}_i \hat{\xi}_i \hat{\xi}_i^{\top} = (v_{ij})_{d\times d}$, is a residual covariance matrix which needs to be sparse. The covariance estimator $\widehat{\Sigma}_{\hat{r}}^{S}$ is defined as:

$$\widehat{\Sigma}_{\hat{r}}^{S} = \sum_{i=1}^{\hat{r}} \widehat{\lambda}_i \widehat{\xi}_i \widehat{\xi_i^{\top}}. \tag{3.28}$$

To make the covariance matrix $\hat{\Gamma}_{\hat{r}}$ sparse, one can use a tresholding technique such as soft-, hard-, adaptive or block thresholding (see for example Ait-Sahalia and Xiu 2017,Fan et al. 2016, Cai and Liu 2011, Jian et al. 2018, Rothman et al. 2008). Ait-Sahalia and Xiu (2017) propose to block-diagonalize $\hat{\Gamma}_{\hat{r}}$ instead of using the soft-or hard thresholding.

In order to use the estimator $\bar{\Sigma}^{S}$, one requires the number of common factors $\hat{r}$. Ait-Sahalia and Xiu (2017) propose an estimator for number of common factors in a large-dimensional dataset. Their estimator is based on PCA of continuous-time models. In this work, we consider the Ait-Sahalia and Xiu estimator to determine the number of factors r. This estimator is defined as:

$$\hat{r} = \underset{1\leq j\leq r_{max}}{\arg\max} \Big(f(j, \hat{\lambda}_j(\hat{\Sigma})) \Big) - 1, \tag{3.29}$$

where r_{max} is some upper bound of $r+1$ and $f(j, \hat{\lambda}_j(\hat{\Sigma}))$ are penalized eigenvalues. The choice of r_{max} does not play any role when estimating the number of common factor, but it justifies an economically significant estimate of $\hat{r}$ in finite sample (see, Ait-Sahalia and Xiu 2017).

The penalized eigenvalues $f(j, \hat{\Sigma})$are defined by:

$$f(j, \hat{\lambda}_j(\hat{\Sigma})) = d^{-1}\hat{\lambda}_j(\hat{\Sigma}) + j \times g(n,d), \tag{3.30}$$

where g denotes the penalty function to be chosen and $\hat{\lambda_j}$ are the realized eigenvalues of covariance matrix $\hat{\Sigma}$. According to Ait-Sahalia and Xiu (2017), the penalty term $g(T,n)$ needs to satisfy the following two criteria:

1. The penalty term needs to be dominated by the signal, defined by the value $d^{-1}\hat{\lambda}_j(\hat{\Sigma})$ when $1 \leq j \leq r$. Since $d^{-1}\hat{\lambda}_j(\hat{\Sigma})$ is $O_p(1)$ as d increases (see, Ait-Sahalia and Xiu 2017), one should select a penalty that shrinks to 0.

2. The penalty should dominate the estimation error as well as the signal value $d^{-1}\hat{\lambda}_j(\hat{\Sigma})$ when $r+1 \leq j \leq d$ to avoid overshooting.

Ait-Sahalia and Xiu (2017) suggest that the penalty term g can be chosen as follows:

$$g(n,d) = \mu\left(\left(\frac{\log d}{n}\right)^{\frac{1}{2}} + 1\right)^{\kappa}, \tag{3.31}$$

where μ and κ are some tuning parameters and $0 < \kappa < 1$. As argued by Ait-Sahalia and Xiu (2017), one can arbitrary choose the tuning parameters, since the covariance matrix estimates are not too sensitive to the number of common factor.

3.4 Monte Carlo Study

3.4.1 Assessing the effects of microstructures noise and non-synchronous trading

In the introduction we highlighted how microstructure effects and non-synchronous trading can jeopardize the estimation of the realized covariance. In this section, we perform a Monte Carlo experiment to asses the performance of the proposed estimator in the presence of these effects. We assume a discretization of 1 second and a trading day of 6.5 hours. This corresponds to $n = 23400$ timestamps. The number of assets d and of common factors r are set to be equal to 10 and 3, respectively. For a given choice of R, Q and H, we generate the observed log-prices through the model in Eq. (3.17). To generate H, we choose H as defined in Eq. (3.18) with unrestricted matrix H^* generates from normal distribution with mean zero and variance one (N(0,1)). As stated above, the covariance matrix R is diagonal and its variances are uniformly generated from [0, 1]. Let Λ_{ii} and R_{jj}, where $i = 1,\dots r; j = 1,\cdots r,\dots d$, denote the diagonal elements of Λ and R, respectively. The diagonal elements of the covariance matrix Q are defined as:

$$\Lambda_{ii} = \bar{\delta} \times \frac{1}{d}\sum_{j=1}^{d} R_{jj},$$

where $\bar{\delta}$ is referred to as signal-to-noise ratio. The off-diagonal elements of the covariance matrix Q, that we denote as $\Lambda_{is}, s, i = 1,2,\dots,r$ with $s \neq i$, are sampled from a standard normal distribution. In order to generate realistic noise scenario, we choose $\bar{\delta} = 0.5, 1$ [4]. To simulate the effect of asynchronous trading on the synchronous observed log-prices, we censor the observed prices using the Poisson sampling. We assume that the probability of having a missing values is equal to $p = 0\%, 50\%$. To set the initial values of the parameters for the EM algorithm, we shift by 50% each element of covariance

[4] These values are observed in real Market, see Buccheri et al. (2019)

matrices R, Λ and factor loading H^* and used the shifted covariance matrices and factor loading as the start point for the EM procedure. We simulate and estimate the model in Eq. (3.2) $N = 1000$ times. For $k = 1, \dots, N$, we recover the parameters $\widehat{R}^k_{ij}, \widehat{\Lambda}^k_{ij}, \widehat{H}^k_{ij}$. We consider the pivotal statistics $\Phi^k_{R_{ij}} = R_{ij} - \widehat{R}^k_{ij}$, $\Phi^k_{\Lambda_{ij}} = \Lambda_{ij} - \widehat{\Lambda}^k_{ij}$ and $\Phi^k_{H_{ij}} = H_{ij} - \widehat{H}^k_{ij}$, where Λ, R and H are the true parameters used to simulate the log-prices. We normalize each of the pivotal statistic $\Phi_{m_{ij}}$, where $m_{ij} \in \{R_{ij}, \Lambda_{ij}, H_{ij}\}$ by their samples of standard deviation. The two scenarios $p = 0\%, 50\%$ are combined with the two scenarios $\bar{\delta} = 0.5, 1$, and therefore we have in total 4 scenarios. We perform the one-sample $t-$test of each element of three matrices $\Phi^k_{R_{ij}}, \Phi^k_{H_{ij}}$ and $\Phi^k_{\Lambda_{ij}}$ to test the null assumption that the mean is zero.

In Table (**??**), we report the sample mean and standard deviation of each element of three matrices Φ^k_R, Φ^k_H and Φ^k_Λ as well as theirs $p-$values obtained with the one-sample $t-$test. We observe that the distribution of Φ^k_R, Φ^k_H and Φ^k_Λ is always centred in 0. This implies that $\hat{R}$, $\hat{H}$ and $\hat{\Lambda}$ correctly estimate the true matrices of R, Λ and H. Figures (3.2) and (3.1) plot the kernel density estimates of the normalized pivotal statistics of $\{\Phi_m\}_{m \in \{R,\Lambda,H\}}$ computed in both scenarios $p = 0, \bar{\delta} = 0.5$ and $p = 0.5, \bar{\delta} = 0.5$. Let's first consider the case in which there are no missing values ($p = 0, \bar{\delta} = 0.5$). From figures 3.1(a), (b) and 3.2(a), (b), (c), (d) we observe that the distributions of the normalized pivotal statistics of Φ_Λ, Φ_H and Φ_R are consistent with the normal distribution. Figures 3.1(b), 3.2(b) and 3.2(d) show, in the scenario $\bar{\delta} = 0, p = 0.5$, the distributions of normalized pivotal statistics of Φ_H, Φ_Λ and Φ_R. They remain consistent with the normal distribution even when half (on average) of the observed prices are removed. The results suggest that the estimators of $\hat{R}$, $\hat{H}$ and $\widehat{\Lambda}$ provide consistent estimates of the true matrices of R, Λ and H even in the case that the data are asynchronized.

These results are not surprising, as we know from the theory of linear-Gaussian models, that maximum likelihood estimates are consistent and asymptotically normal distributed. It is more interesting to examine the case in which there are misspecification, like time-varying covariance matrices. We study this case in next section.

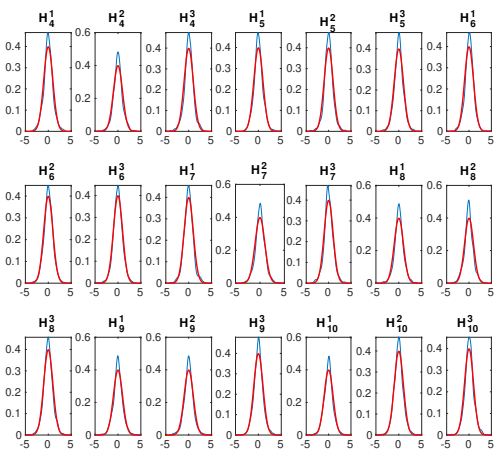
3.4.2 Robustness to stochastic volatility

In the previous section, the covariance matrix Λ has been assumed to be constant over time. However, real high-frequency data are characterized by relevant changes in their covariance structure during the day, and assuming that Λ is constant might be too restrictive. Here, we want to asses the properties of the estimators $\widehat{R}, \widehat{\Lambda}$ and $\widehat{H}$ when the DGP is generated by a misspecified model. In line with the result of Shephard and Xiu (2017), who proved a similar result for the realized covariance matrix recovered from the QMLE, we expect that even the realized factor QMLE is consistent and asymptotically normal distributed.

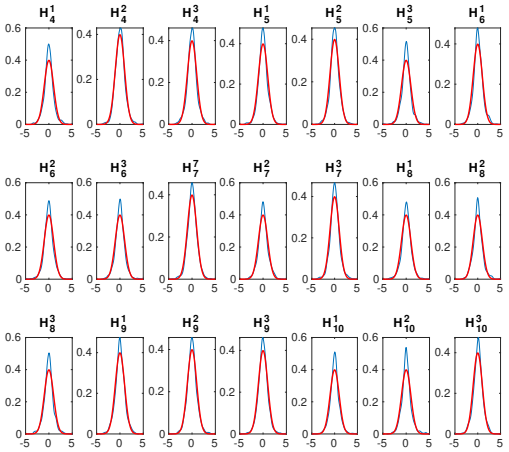
	Mn×100	StDev×100	$p-$value	Mn×100	StDev×100	$p-$value	Mn×100	StDev×100	$p-$value	Mn×100	StDev×100	$p-$value
		$\delta = 0.5$			$\delta = 1$			$\delta = 0.5$			$\delta = 1$	
			$p = 0.5$						$p = 0$			
R_1^1	0.0165	3.4779	**0.8807**	-0.0641	5.1942	**0.6963**	-0.0758	2.1053	**0.2553**	-0.0476	3.0484	**0.6215**
R_2^2	-0.0902	2.9779	**0.3384**	-0.0997	4.8347	**0.5143**	0.0220	1.8883	**0.7122**	0.1195	2.8560	**0.1860**
R_3^3	0.0257	3.5690	**0.8196**	0.1029	5.4641	**0.5517**	-0.0241	1.9888	**0.7013**	0.0164	3.0680	**0.8660**
R_4^4	-0.0083	1.0613	**0.8048**	-0.0263	1.0580	**0.4329**	-0.0655	0.7662	0.0069	0.0126	0.7706	**0.6055**
R_5^5	-0.0017	1.1915	**0.9631**	-0.0829	1.2553	0.0370	0.0032	0.8766	**0.9072**	0.0298	0.8530	**0.2694**
R_6^6	-0.0231	1.2949	**0.5731**	-0.1056	1.2291	0.0067	-0.0466	0.9078	**0.1049**	-0.0402	0.8724	**0.1452**
R_7^7	-0.0541	1.2851	**0.1835**	0.0313	1.2846	**0.4417**	0.0169	0.8728	**0.5407**	0.0177	0.9318	**0.5478**
R_8^8	-0.0352	0.9684	**0.2506**	-0.0408	0.9459	**0.1725**	-0.0187	0.6679	**0.3773**	-0.0652	0.6783	0.0024
R_9^9	-0.0018	1.0811	**0.9581**	0.0098	1.1289	**0.7844**	-0.0221	0.7852	**0.3743**	-0.0174	0.7918	**0.4885**
R_{10}^{10}	-0.0599	1.0872	**0.0819**	-0.0178	1.0581	**0.5946**	-0.0245	0.7399	**0.2962**	-0.0296	0.7519	**0.2136**
Q_1^1	-0.0510	3.7006	**0.6632**	0.1611	6.5854	**0.4395**	0.2043	2.8870	0.0255	5.4168	0.0889	**0.6038**
Q_1^2	2.9885×10^{-4}	2.2435	**0.9966**	0.2418	3.8648	0.0482	0.0065	1.6293	**0.8989**	0.0053	2.6382	**0.9491**
Q_1^3	0.0451	2.3139	**0.5374**	0.1203	4.2772	**0.3741**	0.0101	1.5935	**0.8405**	-0.0648	2.7162	**0.4506**
Q_2^1	2.9885×10^{-4}	2.2435	**0.9966**	0.2418	3.8648	0.0482	0.0065	1.6293	**0.8989**	0.0053	2.6382	**0.9491**
Q_2^2	0.0584	3.3894	**0.5857**	0.1942	6.7041	**0.3598**	0.0425	2.7743	**0.6286**	-0.1624	5.2117	**0.3245**
Q_2^3	-0.0526	2.2213	**0.4542**	-0.1182	3.8816	**0.3358**	0.0404	1.5315	**0.4040**	-0.0123	2.6144	**0.8816**
Q_3^1	0.0451	2.3139	**0.5374**	0.1203	4.2772	**0.3741**	0.0101	1.5935	**0.8405**	-0.0648	2.7162	**0.4506**
Q_3^2	-0.0526	2.2213	**0.4542**	-0.1182	3.8816	**0.3358**	0.0404	1.5315	**0.4040**	-0.0123	2.6144	**0.8816**
Q_3^3	-0.1141	3.5097	**0.3042**	-0.3147	6.6146	**0.1327**	-0.1288	2.7819	**0.1434**	-0.0564	5.3982	**0.7410**
	$(\times 10^{-4})$			$(\times 10^{-4})$			$(\times 10^{-4})$			$(\times 10^{-4})$		
Λ_4^1	-0.30771	0.0168	**0.9537**	-2.0218	0.0106	**0.0106**	-0.38306	0.0118	**0.9180**	-0.70237	0.0081	**0.7827**
Λ_4^2	-1.7798	0.0167	**0.7360**	-0.24576	0.0112	**0.9446**	3.7846	0.0114	**0.2946**	1.3436	0.0084	**0.6125**
Λ_4^3	3.4454	0.0176	**0.5365**	3.5319	0.0122	**0.3592**	-1.5405	0.0123	**0.6923**	-0.58595	0.0085	**0.8278**
Λ_5^1	3.3998	0.0179	**0.5486**	-8.2540	0.0127	0.0397	1.6286	0.0125	**0.6800**	-0.98843	0.0090	**0.7295**
Λ_5^2	2.1419	0.0178	**0.7029**	-3.2644	0.0134	**0.4417**	2.3836	0.0124	**0.5436**	-3.1852	0.0090	**0.2611**
Λ_5^3	0.94478	0.0178	**0.8665**	-1.5821	0.0125	**0.6885**	3.8212	0.0125	**0.3358**	-5.6922	0.0092	**0.0516**
Λ_6^1	11	0.0182	0.0499	4.9225	0.0122	**0.2022**	9.8334	0.0130	0.0168	2.4139	0.0086	**0.3738**
Λ_6^2	4.7195	0.0176	**0.3956**	-0.18649	0.0129	**0.9636**	-4.5315	0.0124	**0.2475**	2.2325	0.0090	**0.4309**
Λ_6^3	-2.8996	0.0170	**0.5894**	7.5049	0.0128	**0.0643**	1.1178	0.0129	**0.7834**	2.0283	0.0091	**0.4823**
Λ_7^1	-6.6913	0.0181	**0.2425**	3.2211	0.0123	**0.4079**	6.2485	0.0127	**0.1211**	-0.11739	0.0089	**0.9668**
Λ_7^2	7.4515	0.0178	**0.1860**	-0.32994	0.0124	**0.9329**	4.8915	0.0122	**0.2065**	3.7180	0.0086	**0.1723**
Λ_7^3	1.9881	0.0182	**0.7304**	3.7566	0.0137	**0.3854**	4.4551	0.0132	**0.2875**	1.9403	0.0092	**0.5044**
Λ_8^1	9.2030	0.0165	**0.0774**	5.3868	0.0109	**0.1198**	5.8181	0.0113	**0.1052**	1.6136	0.0074	**0.4919**
Λ_8^2	-12	0.0167	0.0207	4.0144	0.0104	**0.2207**	3.6842	0.0108	**0.2830**	-1.6248	0.0076	**0.5008**
Λ_8^3	-5.2775	0.0158	**0.2905**	-3.4611	0.0110	**0.3210**	-4.4369	0.0114	**0.2187**	2.2457	0.0077	**0.3565**
Λ_9^1	0.31547	0.0171	**0.9535**	7.8382	0.0125	0.0481	1.3911	0.0120	**0.7140**	-2.7497	0.0085	**0.3078**
Λ_9^2	1.1099	0.0167	**0.8333**	-3.5533	0.0120	**0.3504**	0.97528	0.0120	**0.7969**	4.7097	0.0084	**0.0765**
Λ_9^3	1.4519	0.0168	**0.7852**	-0.38454	0.0121	**0.9200**	-0.74045	0.0119	**0.8443**	-2.3648	0.0092	**0.4174**
Λ_{10}^1	-3.8655	0.0167	**0.4633**	4.7345	0.0122	**0.2212**	1.7413	0.0122	**0.6511**	-0.17344	0.0088	**0.9505**
Λ_{10}^2	-7.1453	0.0169	**0.1813**	7.5195	0.0118	**0.9752**	5.5564	0.0121	**0.1469**	2.0909	0.0085	**0.4368**
Λ_{10}^3	8.8979	0.0174	**0.1060**	4.3670	0.0119	**0.2458**	8.0159	0.0117	0.0312	0.75102	0.0086	**0.7835**

Notes: The results of mean and standard deviation of all elements of the pivotal matrices $\Phi_R^k, \Phi_H^k, \Phi_\Lambda^k$ and the p-values of the one-sample t-test in all the simulated scenario. The bold numbers report the results in which the null hypothesis is not rejected at the 5% confidence interval.

Figure 3.1 A set of two subfigures: (a) plot of the kernel density estimate of the standardize pivotal statistic Φ_H for the scenario $\bar{\delta} = 0.5, \Xi = 0.5$; (b) plot of the kernel density estimate of the standardize pivotal statistic Φ_H for the scenario $\bar{\delta} = 0, p = 0.5$. All results are based on 1000 independent random replications. The pivotal statistic Φ_H is standardize by their sample standard deviation. The solid red line is the density of standard normal distribution.



(a)



(b)

We simulate the data using the dynamic factor model in Eq. (3.2) with a time-varying covariance matrix Λ_t. Let us consider the following model

$$\begin{cases} Y_t = HG_t + \eta_t & \eta_t \sim \mathcal{N}(0, R) \\ G_t = G_{t-1} + b_t & b_t \sim \mathcal{N}(0, \Lambda_t) \end{cases} \tag{3.32}$$

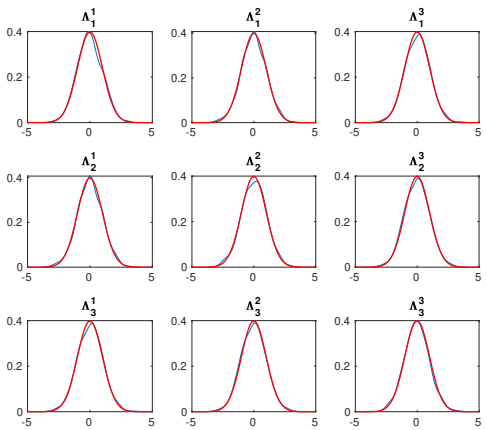
where Λ_t is time varying. We decompose Λ_t as

$$\Lambda_t = D_t C D_t, \tag{3.33}$$

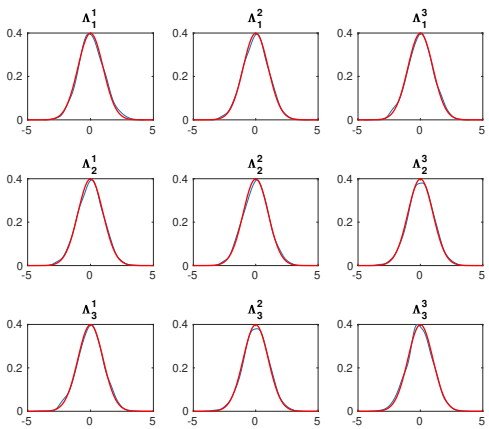
where C is a constant correlation matrix and D_t is a diagonal matrix of a time varying standard deviations

$$D_t = \begin{pmatrix} \sigma_{1,t} & 0 & \dots & 0 \\ 0 & \sigma_{2t} & \dots & 0 \\ \vdots & \vdots & \sigma_{n-1t} & 0 \\ 0 & \cdots & \cdots & \sigma_{nt} \end{pmatrix}. \tag{3.34}$$

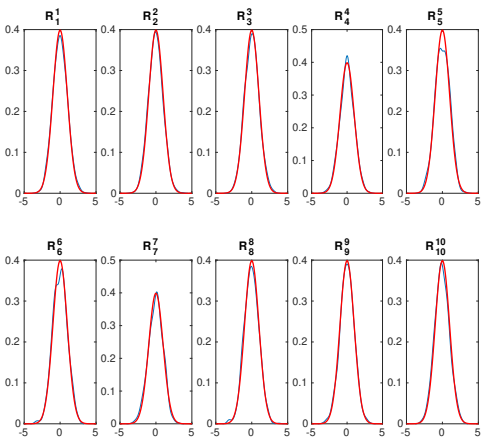
Figure 3.2 A set of four subfigures: (a) plot of the kernel density estimate of the standardize pivotal statistic Φ_Λ for the scenario $\bar{\delta} = 0.5, p = 0.5$; (b) plot of the kernel density estimate of the standardize pivotal statistic Φ_Λ for the scenario $\bar{\delta} = 0.5, p = 0$; (c) plot of the kernel density estimate of the standardize pivotal statistic Φ_Λ for the scenario $\bar{\delta} = 0.5, p = 0.5$; (d) plot of the kernel density estimate of the standardize pivotal statistic Φ_Λ for the scenario $\bar{\delta} = 0, p = 0.5$. All results are based on 1000 independent random replications. The pivotal statistics Φ_R and Φ_Λ are standardize by their sample standard deviation. The solid red line is the density of standard normal distribution.



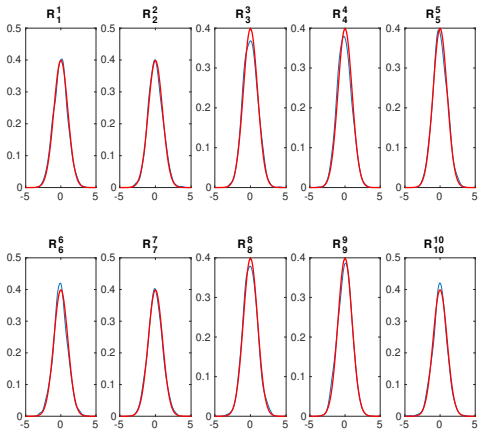
(a)



(b)



(c)



(d)

The diagonal elements of D_t are computed according to the Cox et al. (2005) (CIR) model

$$d\sigma_{jt} = \kappa_j(\theta_j - \sigma_{jt}^2)dt + \upsilon_j \sigma_{jt} dB_{jt}, \quad j = 1, \cdots, r \tag{3.35}$$

where B_j is as standard Brownian motion with $\mathbb{E}[dW_j dB_j] = \rho_j dt$.

We conduct an experiment similar to that in Section (3.4.1) to asses the performance of our estimator under the misspecified DGP. Assume, as before, a discretization of one second. The CIR model is discretized through Euler scheme and we draw the first observation from a Gamma distribution $\Gamma(2\kappa\theta_j/\upsilon_j^2, \upsilon_j^2/2\kappa_j)$ centred in the mean variance. The parameters of the DGP are chosen as follows: $\theta_1 = \theta_2 = 1.0872, \kappa_1 = 20, \kappa_2 = 25, \upsilon = 0.1$. The variances $R_i, i = 1, \ldots, 8$ of the noise are calculated based on the average of signal-noise-ratio $\bar{\delta}_i = \sqrt{\theta_i}/\sum_{i=1}^{d} \text{diag}(R_{ii})/d$. As in the previous Section, we set $\bar{\delta}$ equal to $0.5, 1$. To simulate the effect of asynchronous trading, the observed prices are censored using the Poisson sampling. The probability of having a missing value is equal to $p = 0\%, 50\%$. We fit the misspecified model in Eq.(3.32) to $N = 1000$ independent realizations, and estimate the parameters $\hat{R}_k, \hat{\Lambda}_k$ and $\widehat{H}_k$, with $k = 1, \ldots, N$. The following pivotal statistics are considered: $\Phi_R^k = R - \hat{R}_k, \Phi_\Lambda^k = \hat{\Lambda}_k - \frac{1}{n}\int_0^n \Lambda_s ds$ and $\Phi_H^k = H - \hat{H}_k$. The three pivotal statistic Φ_m, where $m \in \{R, \Lambda, H\}$ are normalized by their sample standard deviation.

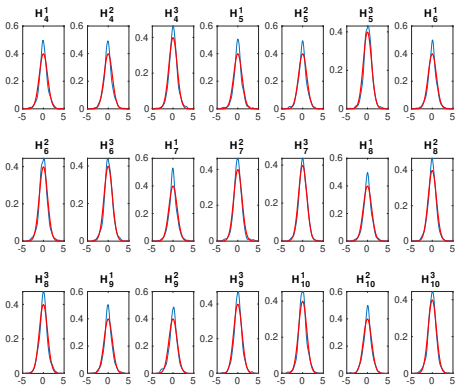
Table (**??**) reports the sample mean and standard deviation of each element of the three matrices Φ_R^k, Φ_H^k and Φ_Λ^k as well as theirs $p-$values obtained with the one-sample $t-$test. The one-sample t-test is performed to test the null assumption that the mean is zero. We observe that as in Section (3.4.1), the distribution of Φ_R^k, Φ_H^k and Φ_Λ^k is always centred in 0. These results suggest that the estimators of $\hat{R}$, $\hat{H}$ and $\hat{\Lambda}$ provide consistent estimates of the true matrices of R, Λ and H even in the case the covariance matrix Λ is time varying. These results are in line with the result in Shephard and Xiu (2017).

Figures 3.3(c), (a), (b) report only the results for the scenario $\Xi = 0.5, \bar{\delta} = 0.5$. We observe that the distributions of the normalized pivotal statistics Φ_Λ, Φ_H and Φ_R are consistent with a normal distribution even when half (on average) of the observed prices are removed. The distributions are always centred around 0, as also shown in Table (**??**).

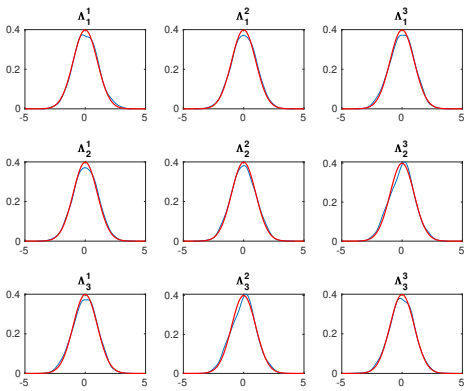
	Mn×100	StDev×100	$p-$value	Mn×100	StDev×100	$p-$value	Mn×100	StDev×100	$p-$value	Mn×100	StDev×100	$p-$value
	$\delta = 0.5$			$\delta = 1$			$\delta = 0.5$			$\delta = 1$		
	$p = 0.5$						$p = 0$					
R_1^1	-0.0414	4.0733	**0.7478**	0.3307	9.7897	**0.2857**	0.1113	2.2287	**0.1145**	-0.2083	5.3934	**0.2221**
R_2^2	-0.0080	3.3854	**0.9405**	-0.7787	9.8529	0.0126	-0.0158	1.9384	**0.7962**	-0.1097	4.7495	**0.4655**
R_3^3	-0.1888	3.8167	**0.1180**	0.4454	9.8719	**0.1539**	0.0912	2.2465	**0.1997**	-0.2926	5.1554	**0.0730**
R_4^4	-0.0424	1.0497	**0.2022**	-0.0122	1.0587	**0.7152**	-0.0157	0.7467	**0.5075**	-0.0425	0.7635	**0.0784**
R_5^5	-0.0391	1.2449	**0.3209**	-0.0160	1.2416	**0.6828**	-0.0304	0.8742	**0.2716**	-0.0016	0.8696	**0.9544**
R_6^6	-0.0078	1.2214	**0.8405**	-0.0202	1.3012	**0.6237**	0.0059	0.8737	**0.8302**	-0.0593	0.8641	0.0302
R_7^7	0.0287	1.2506	**0.4675**	-0.0684	1.3065	**0.0980**	-0.0253	0.8678	**0.3569**	0.0014	0.9350	**0.9622**
R_8^8	-0.0414	0.9024	**0.1476**	8.8518×10^{-4}	0.9002	**0.9752**	-0.0187	0.6679	**0.3773**	-0.0188	0.6530	**0.3625**
R_9^9	0.0172	1.0889	**0.6173**	0.0091	1.1257	**0.7989**	-0.0221	0.7852	**0.3743**	-0.0035	0.7925	**0.8877**
R_{10}^{10}	-0.0439	1.0765	**0.1974**	-0.0444	1.0744	**0.1914**	-0.0078	0.7686	**0.7487**	-0.0400	0.7866	**0.1080**
Q_1^1	0.2773	5.2964	**0.0981**	11.2801	16.3302	**0.1678**	-0.0865	4.0650	**0.5014**	0.8624	12.8772	0.0344
Q_1^2	0.0830	3.3719	**0.4367**	0.2418	3.8648	0.0482	0.0490	2.4305	**0.5237**	0.2827	7.3862	**0.2264**
Q_1^3	-0.0386	3.4310	**0.7223**	0.2939	10.2413	**0.3644**	-0.0580	2.3428	**0.4337**	0.1802	7.3325	**0.4371**
Q_2^1	0.0830	3.3719	**0.4367**	0.4924	11.2801	**0.1678**	0.0490	2.4305	**0.5237**	0.2827	7.3862	**0.2264**
Q_2^2	-0.0357	4.8103	**0.8145**	1.0180	17.1507	**0.0608**	-0.0940	3.9289	**0.4494**	0.3478	12.3953	**0.3751**
Q_2^3	-0.0132	3.2976	**0.8996**	0.0210	10.2671	**0.9485**	-0.0852	2.3606	**0.2538**	0.1930	7.3344	**0.4056**
Q_3^1	-0.0386	3.4310	**0.7223**	0.2939	10.2413	**0.3644**	-0.0580	2.3428	**0.4337**	0.1802	7.3325	**0.4371**
Q_3^2	-0.0132	3.2976	**0.8996**	0.0210	10.2671	**0.9485**	-0.0852	2.3606	**0.2538**	0.1930	7.3344	**0.4056**
Q_3^3	-0.0512	5.0029	**0.7465**	-0.7162	15.9169	**0.1551**	-0.0976	3.9647	**0.4367**	0.6517	12.5076	**0.0997**
	$(\times10^{-4})$			$(\times10^{-4})$			$(\times10^{-4})$			$(\times10^{-4})$		
Λ_4^1	0.8393	0.0174	**0.8790**	6.8194	0.0091	0.0175	1.6742	0.0121	**0.6611**	0.7701	0.0061	**0.6916**
Λ_4^2	-13	0.0173	0.0221	-4.8420	0.0083	**0.0664**	-1.8683	0.0123	**0.6309**	-1.2773	0.0062	**0.5166**
Λ_4^3	9.0836	0.0176	**0.1033**	-6.3323	0.0088	0.0228	-4.1436	0.0113	**0.2450**	1.0629	0.0061	**0.5832**
Λ_5^1	-2.6634	0.0194	**0.6642**	2.1806	0.0090	**0.4449**	5.1455	0.0130	**0.2119**	0.8104	0.0067	**0.7018**
Λ_5^2	-8.9369	0.0183	**0.1220**	1.1912	0.0098	**0.7007**	-1.2031	0.0121	**0.7534**	-3.6423	0.0068	**0.0884**
Λ_5^3	19	0.0184	0.0011	-0.0105	0.0089	**0.9970**	-0.7957	0.0128	**0.8442**	5.5385	0.0066	0.0084
Λ_6^1	-2.6988	0.0182	**0.6399**	2.0004	0.0097	**0.5130**	-2.9782	0.0130	**0.4686**	0.8249	0.0065	**0.6866**
Λ_6^2	-3.7439	0.0191	**0.5357**	-3.8515	0.0094	**0.1934**	3.3734	0.0128	**0.4059**	-1.9005	0.0068	**0.3752**
Λ_6^3	7.2454	0.0183	**0.2120**	1.5787	0.0090	**0.5779**	1.6716	0.0124	**0.6707**	3.0953	0.0065	0.1309
Λ_7^1	-1.2592	0.0189	**0.8333**	-0.1382	0.0092	**0.9620**	3.0757	0.0127	**0.4426**	0.377	0.0065	**0.8551**
Λ_7^2	-5.6599	0.0191	**0.3483**	-3.7951	0.0096	**0.2129**	-2.9286	0.0131	**0.4793**	-1.1477	0.0064	**0.5683**
Λ_7^3	7.4727	0.0184	**0.1985**	1.7322	0.0088	**0.5359**	-1.317	0.0126	**0.7401**	0.9321	0.0063	**0.6380**
Λ_8^1	-1.1756	0.0155	**0.8100**	0.1911	0.0080	**0.9395**	-4.8651	0.0112	**0.8906**	0.6091	0.0054	**0.7217**
Λ_8^2	-6.7190	0.0156	**0.1748**	-1.3577	0.0088	**0.6241**	-2.6546	0.0110	**0.4445**	0.1571	0.0057	**0.9308**
Λ_8^3	0.6571	0.0160	**0.8969**	7.4299	0.0080	0.0034	3.5045	0.0111	**0.3179**	-0.128	0.0056	**0.9421**
Λ_9^1	-3.0588	0.0183	**0.5963**	2.0796	0.0090	**0.4630**	-4.8480	0.0125	**0.2218**	1.2617	0.0060	**0.5049**
Λ_9^2	2.4661	0.0179	**0.6629**	1.4783	0.0095	**0.6232**	1.4301	0.0125	**0.7179**	-0.3719	0.0062	**0.8485**
Λ_9^3	7.1813	0.0176	**0.1966**	-1.2881	0.0094	**0.6641**	-5.2439	0.0126	**0.8957**	0.9201	0.0060	**0.6299**
Λ_{10}^1	1.2184	0.0179	**0.8300**	-5.4789	0.0091	**0.0562**	1.3201	0.0127	**0.7433**	0.5048	0.0063	**0.7974**
Λ_{10}^2	-1.8889	0.0174	**0.7315**	-4.2560	0.0089	**0.1295**	1.8882	0.0121	**0.6217**	-1.5919	0.0064	**0.4336**
Λ_{10}^3	-4.6770	0.0172	**0.3905**	5.7463	0.0087	0.0373	-3.0990	0.0116	**0.4003**	0.0207	0.0062	**0.9915**

Notes: We show the mean and the standard deviation of all the elements of the pivotal matrices $\Phi_R^k, \Phi_H^k, \Phi_\Lambda^k$ and the $p-$values of the one-sample $t-$test in all the simulated scenario. The bold numbers report the results in which the null hypothesis is not rejected at the 5% confidence interval.

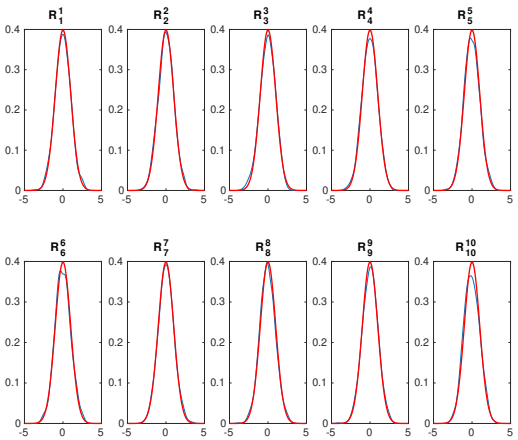
Figure 3.3 A set of three subfigures: (a) plot of the kernel density estimate of the standardize pivotal statistic Φ_H for the scenario $\bar{\delta} = 0.5, p = 0.5$; (b) plot of the kernel density estimate of the standardize pivotal statistic Φ_Q for the scenario $\bar{\delta} = 0.5, p = 0.5$; (c) plot of the kernel density estimate of the standardize pivotal statistic Φ_R for the scenario $\bar{\delta} = 0.5, p = 0.5$. All results are based on 1000 independent random replications. The pivotal statistic Φ_R, Φ_Q and Φ_H are standardized by the sample of his standard deviations. The solid red line is the density of standard normal distribution.



(a)



(b)



(c)

3.4.3 A closer look to PCA-based methods

In order to use the estimator in Eq. (3.29) to extract factors in PCA-based methods, we rely on previous-tick synchronization to reconstruct log-prices because of missing values. A drawback of this approach is that it leads to a large number of

zeros returns, which are known to jeopardize the inference of realized covariances (see e.g. Epps 1979, Hayashi and Yoshida 2005, Buccheri et al. 2018). Since the estimator in Eq. (3.29) relies on the estimation of the realized covariance matrix, the question that arises is whether the PCA-based methods can consistently estimate the number of factors in high-frequency data. Our goal is to study the impact of zeros returns on the estimation of the number of common factor based on PCA-based methods.

The effect of the zeros returns on the estimated number of factors: heuristic

In this section, we provide the intuition on the effect of previous-tick interpolation on the spectral properties of the realized covariance matrix.

- Assume that d is equal to two for simplicity and assume that the *integrated* covariance matrix Q has rank equal to 1. We can therefore write :

$$\begin{aligned} Q &= X * X' \\ &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (x_1, x_2) = \begin{pmatrix} x_1^2 & x_1 x_2 \\ x_1 x_2 & x_2^2 \end{pmatrix}. \end{aligned}$$

  One eigenvalue of Q is zero. To determine determine the other, we need to find the values of λ which satisfy the characteristic equation of the matrix Q, namely those values of λ for which

$$\det(Q - \lambda I_2) = 0.$$

  It is immediate to see that :

$$\lambda_1 = x_1^2 + x_2^2 \quad \& \quad \lambda_2 = 0. \tag{3.36}$$

- Let us now assume that data are asynchronous. We synchronize them through the previous tick-interpolation before estimating the realized covariance. Let RC^* be the realized covariance matrix of the synchronized data. Buccheri et al. (2018) show that

$$RC^* \xrightarrow{p} Q^* = Q \circ A$$

where A is a symmetric matrix with all off-diagonal entries lower than one and diagonal elements equal to 1. The matrix A can be therefore written as:

$$A = \gamma \mathbf{1}_2 \mathbf{1}_2' + (1-\gamma) I_2, \qquad 0 \leq \gamma \leq 1, \tag{3.37}$$

where γ is a bias parameter, $\mathbf{1}$ is a 2×1 column vector of ones and I_2 is the 2×2 identity matrix. The parameter γ describes the effect of zeros returns on the off-diagonal elements of the integrated covariance Q. Thus, we can write:

$$\begin{aligned} A &= \gamma \begin{pmatrix} 1 \\ 1 \end{pmatrix} (1,1) + (1-\gamma) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad 0 \leq \gamma \leq 1 \\ &= \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}. \end{aligned} \tag{3.38}$$

And therefore

$$\begin{aligned} Q^* &= Q \circ A \\ &= \begin{pmatrix} x_1^2 & x_1 x_2 \\ x_1 x_2 & x_2^2 \end{pmatrix} \circ \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix} = \begin{pmatrix} x_1^2 & \gamma x_1 x_2 \\ \gamma x_1 x_2 & x_2^2 \end{pmatrix}. \end{aligned} \tag{3.39}$$

The eigenvalues of Q^* are given by

$$\begin{aligned} \lambda_1^* &= \frac{x_1^2 + x_2^2 + \sqrt{(x_1^2 + x_2^2)^2 - 4(1-\gamma^2) x_1^2 x_2^2}}{2} \\ \lambda_2^* &= \frac{x_1^2 + x_2^2 - \sqrt{(x_1^2 + x_2^2)^2 - 4(1-\gamma^2) x_1^2 x_2^2}}{2}. \end{aligned} \tag{3.40}$$

If $\gamma = 1$, we observe that the eigenvalues of Q^* are equal to those of Q. This is not surprising, because when λ is equal to one, Q^* reduces to Q. If $0 \leq \gamma < 1$, the eigenvalues of Q^* are different from those of Q. In particular, while Q is reduced rank, we recover a full rank matrix Q^*. Thus, the spectral properties of the integrated covariances are significantly affected by zero returns.

The affect of the zero returns on the estimation of the number of factors: simulation evidence

In this section, we examine through the simulations the above findings. In particular, we examine whether zero returns affect the properties of the estimator in Eq. (3.29), as the latter relies on the estimated realized covariance matrix.

We generate the data through the model in Eq. (3.32)-(3.35) for $T = 23400$. We set the number of assets d and factors r to be equal to 100 and 5, respectively. We discretize CIR model through the Euler scheme and draw the first observation from a Gamma distribution $\Gamma(2\kappa\theta_j/\upsilon_j^2, \upsilon_j^2/2\kappa_j)$ centred in the mean variance. The parameters of CIR model (3.35) are set as: $\theta_1 =, \cdots = \theta_5 = 0.03, \kappa_1 =, \cdots = \kappa_5 = 5; \upsilon_1 = \cdots = \upsilon_5 = 0.1$. The variances $R_i, i = 1, \ldots, 100$ of noise are calculated based on the average of signal-noise ratio

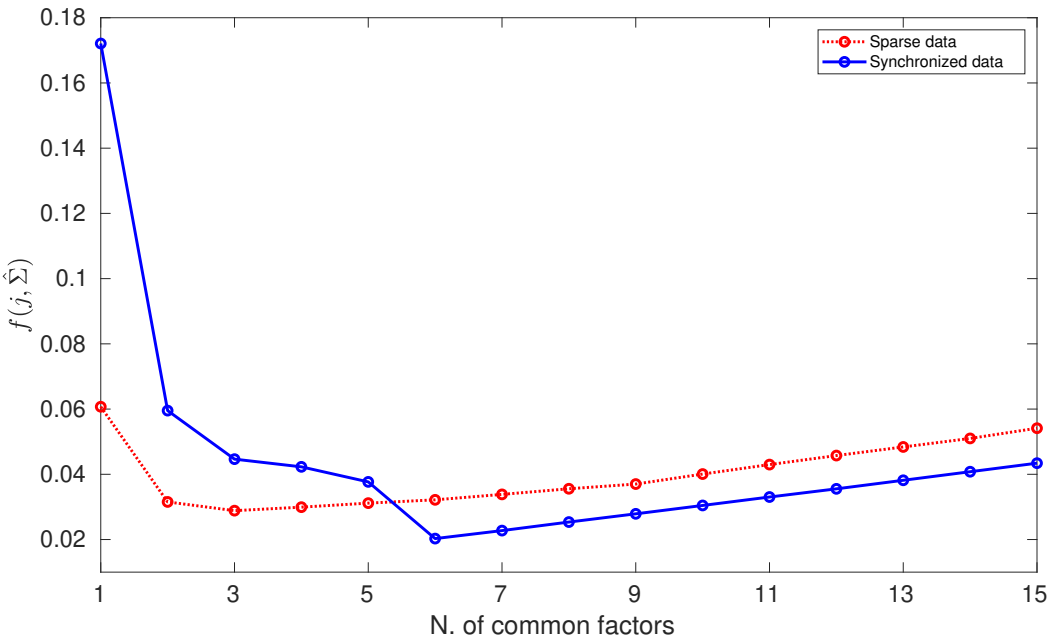
$$\bar{\delta}_i = \sqrt{\theta_i}/((\sum_{i=1}^{d} \text{diag}(R_{ii}))/d).$$

We choose the average signal-noise ratio as $\bar{\delta} = 0.1$. To mimic the effect of asynchronicity in the simulated log-prices, we censor the observed prices using Poisson sampling. We assume that the probability of having a missing value is equal to $p = 50\%$. We choose the tuning parameters of the penalty function g of the estimator in Eq. (3.29) as $\kappa = 0.5, r_{\max} = 20$ and $\mu = 0.05 \times \hat{\lambda}_{\min(\frac{n}{2}, \frac{T}{2})}$. The median eigenvalues $\hat{\lambda}_{\min(\frac{n}{2}, \frac{T}{2})}$ is used to adjust the level of the average eigenvalues in order to improve the accuracy of the estimator (Ait-Sahalia and Xiu 2017). The Monte Carlo experiments are based on 1000 replications. In each replication, we sample the observations at one minute in order to reduce the affect of both the microstructure noise and the zero returns on the PCA-based methods. This procedure is commonly used in the computation, for instance, of the realized covariance. The disadvantage of sampling at a lower frequency is that much of the data is largely neglected. However, sampling at a lower frequency helps to improve the estimation of common factors with the PCA-based methods.

We start our analysis by first comparing the penalized eigenvalues $f(j, \hat{\Sigma})$ and the penalized eigenvalues $f(j, \bar{\Sigma})$. The realized covariance $\hat{\Sigma}$ is estimated on synchronized data, while $\bar{\Sigma}$ is estimated on non-synchronized data. The penalized eigenvalue $f(j, \cdot)$ is given in Eq. (3.30).

Figure 3.4 shows a comparison between the penalized eigenvalues of the realized covariance estimated on non-synchronized data ($f(j, \bar{\Sigma})$) and the penalized eigenvalues of the realized covariance matrix estimated on synchronized data ($f(j, \hat{\Sigma})$). First, we observe that the penalized eigenvalues $f(j, \hat{\Sigma})$ are significantly different from $f(j, \bar{\Sigma})$, confirming what we observed in Section (3.4.3). In particular, we observe a significant difference between the minimum of $f(j, \bar{\Sigma})$ and that of $f(j, \hat{\Sigma})$. In fact, the minimum value of $f(j, \bar{\Sigma})$ is of 3, whereas that of $f(j, \hat{\Sigma})$ is of 6. This difference is due to the effect of zero returns on the realized covariance matrix.

Figure 3.4 The average penalized eigenvalues of realized covariance matrix estimate with synchronous simulated log-prices and the penalized eigenvalues of realized covariance matrix estimate with asynchronous simulated log-prices.



Notes: This figure reports the results of the penalized eigenvalues of realized covriance matrix estimate with synchronized simulated data ($f(j,\hat{\Sigma})$) and the penalized eigenvalues of realized covriance matrix estimates on non-synchronized simulate data ($f(j,\bar{\Sigma})$). $\hat{\Sigma}$ is realized covariance matrix estimates on synchronized simulate data, whereas $\bar{\Sigma}$ is the realized covariance estimates on non-synchronized simulate data.

Table 3.1 presents the result of the average number of factors selected across 1000 replications. As expected, the PCA-based methods consistently estimates the true number of common factors, since average of the number of common factors selected across the replications is around 5. On the other hand, when the simulated data contain missing observations, the PCA-based methods does not consistently recover the true number of factor. We observe that the average number of common factors selected across the replications is 2, which is actually very small compared to the true number of factor 5. This result shows that the PCA-based methods appears to underestimate the true number of factors.

Table 3.1 Summary statistics for the average number of factors over 1000 replications.

Avg. estimated r	Synchronized data	Non-synchronous data
$\bar{\hat{r}}$	5	2

Notes: This table reports the average over 1000 replications of the number of common factors estimated with simulate synchronous log-prices and non-synchronous simulated log-prices.

3.5 Empirical application

The purpose of this section is to asses the performance of the proposed estimator when applied to real high-frequency data. We first perform the AIC, BIC and PCA-based methods on real high-frequency data to determine the number of common factors. We compare the number of factors determined by the AIC and BIC criteria with that determined by PCA-based methods. We then asses the performance of the realized covariance in on out-of-sample global minimum variance portfolio and in Value-at-Risk forecasting.

3.5.1 The Dataset

The dataset under consideration is provided by Thomson Reuters and contains intraday transaction data of the 100 most liquid assets traded in the NYSE. The data range from January 3*rd*, 2006 to December 31*st*, 2014, and thus the total number of trading days is 2265. The exchange opens at 9.30 and closes at 16.00 local time. The time resolution is one second, so the maximum daily sample size is 23.400. The data is cleaned through the procedure described by Barndorff-Nielsen et al. (2009).

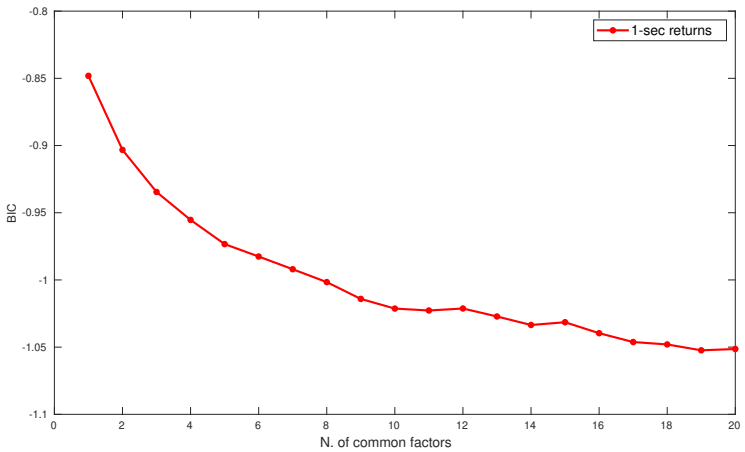
3.5.2 In-sample analysis and diagnostic

To perform the AIC and BIC tests on high-frequency data, we randomly select 15 trading days. For each criterion (AIC and BIC), the estimate of the number of factors r is set to be equal to the number that minimizes the criterion from a potential number of factors between 1 and 20. It is well-known that the AIC criterion does appear to overestimate the number of common factors, and thus we will only focus on the results of BIC criterion.

Figure 3.5 shows the result of BIC criterion. BIC are averaged over the random sample of 15 days. We observe a slow decay of the BIC and, therefore, the BIC criterion selects no factor among the potential number factors between 1 and 20.

T

Figure 3.5 Normalized BIC of daily returns and of 1-second returns. In the second case, we plot the average BIC over a random sample of 15 days.



To asses whether the slow decay of the BIC is due to model misspecification, e.g. stochastic volatility, assume that the number of common factor is equal to $5, 10, 15, 20$ and 100, and thus we focus on the following models:

- QMLE_5 : realized factor QMLE, with $r = 5$
- QMLE_{10} : realized factor QMLE, with $r = 10$
- QMLE_{15} : realized factor QMLE, with $r = 15$
- QMLE_{20} : realized factor QMLE, with $r = 20$
- QMLE : realized factor QMLE, with $r = d$

The case $r = d$ coincides with the QMLE of Corsi et al. (2015) and Shephard and Xiu (2017). For the sake of simplicity, we perform this analysis in the time period from January 2011 to December 2012. We estimate the dynamic factor models on each trading day of 2011-2012, and perform standard diagnostic tests on linear Gaussian state-space models. Denote by $j = 1, 2, \ldots, 502$, the trading days and by $t = 1, 2, \ldots, 23400$ intraday timestamps. The prediction error of the QMLE$_i$, with $i = 5, 10, 15, 20, 100$, denoted by E_j^i is computed as:

$$E_j^i(k) = \left(Y_j(k) - \widehat{H}_j^i(k)(\widehat{G}_t^{t-1})_j^i \right)^2, \tag{3.41}$$

where $k = 1, \ldots, 100$, $(\widehat{G}_t^{t-1})_j^i$ are filtered factors given by equation (B.1). $Y_j(k)$ is the log-price of the asset k at day j and $\widehat{H}_j^i$ is the estimated factor loading matrix of model i at day j. In order to compute $(\widehat{G}_t^{t-1})_j^i$, we use the set of the estimated parameters $\widehat{\Omega}_{ij}$[5] of the QMLE $_i$. The mean square error (MSE) can be computed by averaging the prediction error (3.41) over $j = 1, \ldots, 501$; i.e.

$$V_t^i(k) = \frac{1}{N} \sum_{j=1}^{N} E_j^i(k), \tag{3.42}$$

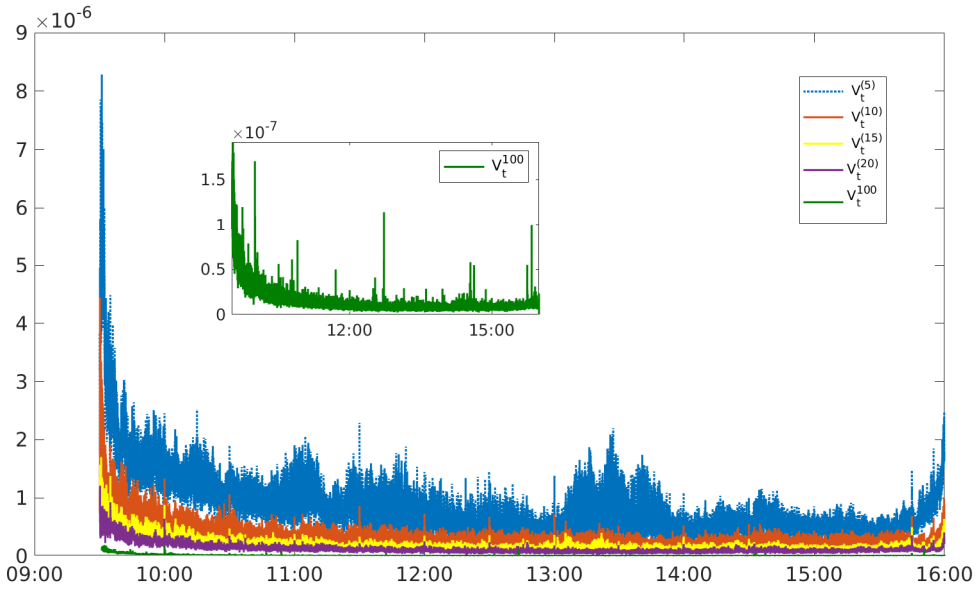
where $N = 502$.

Figure 3.6 shows, for the first asset, the MSE obtained with the dynamic factor models. First, we observe that the MSE provided by the QMLE$_5$, QMLE$_{10}$, QMLE$_{15}$ and QMLE$_{20}$ exhibits the well known U-shape. The MSE is large at the beginning of the day (from 9h30 to 11h), and it slowly decreases until at the last few minutes, when it sharply increases. The large values of the MSE at the beginning of day are due to idiosyncratic trading, reflecting the uncertainty coming from the overnight market and the opening price. The QMLE$_5$ has largest MSE, whereas the QMLE has lowest MSE. This is not surprising, as the QMLE has a large number of parameters and fits data better in-sample. Looking at the figure inside figure 3.6, we observe that even if the QMLE provides the lowest MSE, it is also affected by idiosyncratic trading at the beginning of the trading day. However, this is not an issue from the point of view of RC estimation because of the result of Shephard and Xiu (2017).

We saw that the BIC has a slow decay on intraday data. On the other hand, we saw that idiosyncratic trading represents an important source of misspecification for the model, as it leads to heteroskedastic prediction errors. The question that arises is: does idiosyncratic trading affect the factor structure? To understand whether the idiosyncratic trading affects the factor structure, we perform the BIC test on the sub-sample obtained by removing data from 9:30 to 11:00 and from 15:30

[5] $\hat{\Omega} = \{\widehat{H}, \widehat{\Lambda}_r, \widehat{R}\}$.

Figure 3.6 Average intraday volatility provides by the realized factor QMLE and QMLE. Insert: zooms the time series of the average intraday volatility obtained with QMLE.



to 4:00 (illustrated in figure 3.7). In other words, we remove data on the first and last trading hours to mitigate the effect of idiosyncratic trading.

Figure 3.7 The description of the full-sample data and sub-sample data

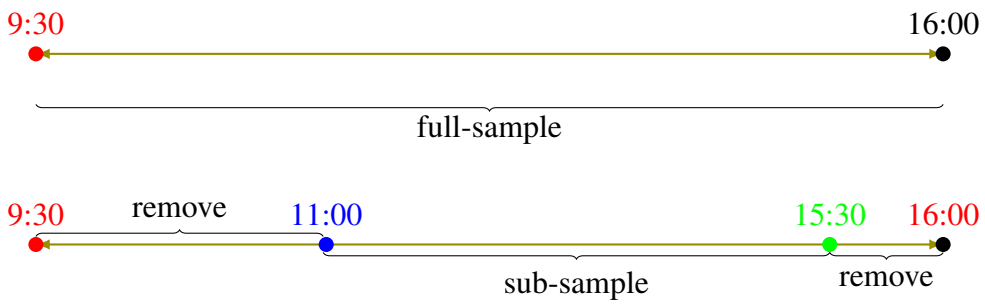
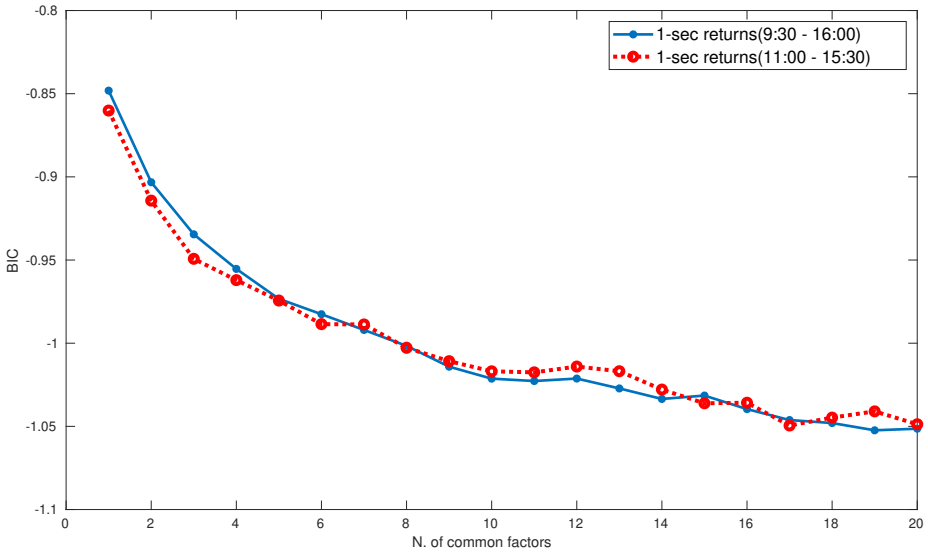


Figure 3.8 reports the results of the BIC criterion obtained in the full-sample and in the sub-sample. The BIC results obtained in the full-sample and in the sub-sample are averaged over the 15 trading days. We observe that the results of the BIC obtained in the sub-sample are similar to those obtained in the full-sample across all the numbers of common factors. This similarity indicates that idiosyncratic trading does not affect the factor structure.

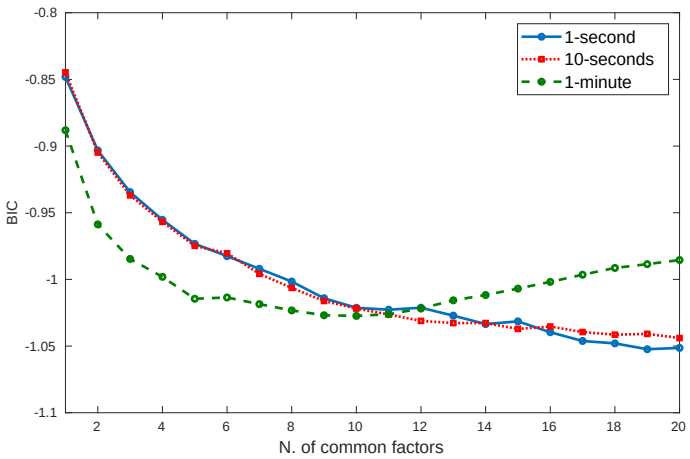
Figure 3.8 We show the normalized average full-sample BIC and normalized average sub-sample-BIC.



3.5.3 The role of the sampling frequency

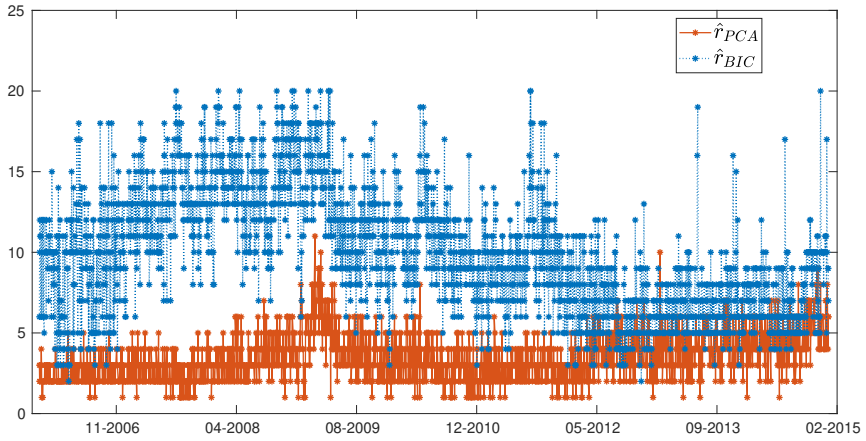
Ultra-high frequency data are characterized by strong lead-lad dependencies that may affect the estimation of the dynamic factor model in Eq. (3.2). Indeed, we are only focusing on contemporaneous dependencies and thus potential lead-lag dependencies are neglected by our model. In order to mitigate the effect of the lead-lag dependencies, we aggregate the prices at a lower frequency. First, we sample at 10-second each of the 15 randomly selected trading days. We report the results in Fig. 3.9. Note that even when prices are sampled at 10-seconds frequency, the BIC has a similar decay. Then, we sample at 1-minute and average the recovered BIC over the 15 randomly selected trading days. In contrast to the 1-second and 10-second prices, the BIC of 1-minute prices becomes flat at around $r = 4$ and starts to increase at $r = 10$. This indicates that a factor structure emerges at the one minute sampling frequency, where the effects of lead-lag dependencies are weaker.

Figure 3.9 We show the normalized BIC of 1-second sampled, 10-second sampled and 1-minute sampled prices. The BIC results are averaged over the whole sample of 15 trading days.



Next, we estimate the number of factors on the entire sample of 2265 trading days using both PCA and BIC. To this end, we sample at 1-minute each of the 2265 trading days. Figure 3.10 shows the time series of the estimated number of factors for the entire sample obtained with the estimator in Eq. (3.29) and the BIC criterion. Two main findings emerge from this result. First, we observe that the number of factors estimated by the BIC criterion, as well as the PCA-based methods, is significantly higher in the time period of 2007-2009. This indicates that both criteria appear to estimate a higher number of common factors during extreme events, such as the global recession in 2009. Second, we observe that the number of factors obtained with the BIC criterion is significantly higher than that obtained with the PCA based methods for almost all days. The BIC criterion identifies most of the number of factor between 5 and 15, whereas the PCA estimator identifies most of the number of factors between 3 and 5. These differences are mainly due to the effect of asynchronous trading and zero returns on the PCA-based methods. Indeed, as showed in Section (3.4.3), the PCA-based methods underestimate the true factors when the sample data are interpolated with previous-tick.

Figure 3.10 We show the time series of the number of common factors estimated by the BIC and PCA-based method on 1-min sampled date for the entire sample of 502 trading days.



In addition, Table 3.2 presents the summary statistic and distributions of the time series of $\hat{r}_{\text{BIC}}$ and $\hat{r}_{\text{PCA}}$. First, we observe that the mean and median of the estimated common factors with the BIC criterion are 9.95 and 10, respectively. We also observe that 13% of the estimated common factors are almost below 5 factors, while 91% do not exceed the 15 factors. Second, we observe that the mean and median of the estimated common factors with the PCA-based methods are 3.5 and 3, respectively. Furthermore, the 90% of the estimated common factors are almost below 5 factors, while 100% do not exceed the 15 factors. The PCA-based method seems to underestimate the number of common factor compared to the dynamic factor model. In the next Section, we examine the consequences of these differences in out-of-sample portfolio construction and VaR assessment.

Table 3.2 The summary statistics for the estimated number of common factors

	Mean	Median	The case in which the estimated common factors are			
			$\hat{r}_{\text{BIC}} \leq 5$	$\hat{r}_{\text{BIC}} \leq 10$	$\hat{r}_{\text{BIC}} \leq 15$	$\hat{r}_{\text{BIC}} > 16$
$\hat{r}_{\text{BIC}}$	9.95	10	12.67%	59.25%	90.95%	6.137%
$\hat{r}_{\text{PCA}}$	3.5	3	90.02%	99.96%	100%	0%

Notes: This table presents the summary statistic of the number of common factors estimated by the BIC criterion and the PCA-based methods on 1-minute sampled prices for the entire sample of 2265 trading days.

3.5.4 Out-of-sample portfolio construction

We now asses the performance of the proposed covariance estimator through a global minimum variance portfolio (GMV) exercise. We follow the logic of Patton and Sheppard 2009 and Engle and Colacito 2006. Let us assume there are d risky assets on the financial market. We denote by Σ_j the true integrated covariance matrix of the assets returns at day j. We recall that the global minimum variance (GMV) portfolio is the unique solution of the following optimization problem:

$$\widehat{\omega}_j = \min_{\omega_j} \omega_j' \hat{\Sigma}_j \omega, \quad \text{subject to} \quad i_d \omega_j = 1, \tag{3.43}$$

where ω_j are the portfolio weights and i_d is a column vector of ones. When the optimal portfolios $\hat{\omega}$ are obtained, the related (ex-post) variance is given by:

$$EV = ||\widehat{\omega}_j' \Delta Y_j \Delta Y_j' \widehat{\omega}_j|| \tag{3.44}$$

where $||\cdot||$ denotes the Euclidian norm, ΔY is the vector of daily open-to-close returns and $\widehat{\omega}$ is the solution of the problem (3.44). The best covariance estimator is the one providing the lowest ex-post variance. In order to test whether the ex-post volatility of two methods are statistically distinguishable, we use the model confidence set (MCS) of Hansen et al. (2011).In particular, we consider a model confidence set at the 90% confidence level, and it is denoted by $\mathrm{M}_{90\%}$.
For each day from January 3rd, 2006 to December 31st, 2014, we build the daily portfolio as follows: at the end of each day, we compute the realized covariance (RC) matrix with different estimators. We then use the RC matrix as the predictor for the next period and solve the problem in Eq. (3.44). We compute seven different RC estimators.

- The first five estimators are QMLE_5, QMLE_{10}, QMLE_{15}, QMLE_{20} and $\mathrm{QMLE}_{\hat{r}_{bic}}$. The QMLE_5, QMLE_{10}, QMLE_{15} and QMLE_{20} are given by the realized factor QMLE in Eq. (3.2) with the number of factors r equal to $5, 10, 15$ and 20, respectively (see Section 3.5.2). On the other hand, the $\mathrm{QMLE}_{\hat{r}_{bic}}$ is the realized factor QMLE in Eq. (3.2) with the numbers of factor estimated by BIC on each of the 2265 trading days (see Figure 3.10). The RC estimate is given by:

$$\hat{\Sigma}_r = \hat{H}_r \hat{\Lambda}_r \hat{H}_r^\top \tag{3.45}$$

  where $\hat{H}_r$ and $\hat{\Lambda}_r$ are the EM estimates of the factor loadings and of the factor covariance.

- As a benchmark, we consider the PCA estimator($\overline{\Sigma}^S$) defined in Eq. (3.28). We also consider the Multivariate Realized Kernel (MRK) of Barndorff-Nielsen et al. (2011). To implement the MRK, we use a Parzen kernel and use the refresh-time to synchronize the data. We finally consider the quasi-maximum likelihood (QMLE) of Corsi et al. (2015) and Shephard and Xiu (2017). The QMLE estimator is similar to the estimator $\hat{\Sigma}_r$ defined in Eq.(3.45) with $d = r; \hat{H}_r = I$.

Table 3.3 reports the results of the analysis for three different time periods: January 2006 - December 2008, January 2009 - December 2011 and January 2012 - December 2014. In the first column, we observe that the QMLE method provides the lowest ex-post volatility, followed by QMLE_{20}, QMLE_{10}, QMLE_{15}, $\text{QMLE}_{\hat{r}_{bic}}$ and QMLE_5. The ex-post volatility of MRK and PCA is substantially higher. The lower ex-post volatility of QMLEs is due to the fact that QMLE methods exploit all the available one-minute data when estimating the covariance matrix. Unlike QMLEs, PCA and MRK rely on synchronization schemes. The PCA is based on the previous-tick interpolation. This procedure leads to a large number of "or artificial" zero returns when estimating the covariance matrix which in turn bias the estimator. The MRK relies on refresh-time synchronization scheme, which discards a lot of data when estimating the covariance matrix. The p-values of the MCS test are indicated in parenthesis. We observe that the p-value of QMLEs is significantly higher, indicating that QMLEs methods perform better than PCA and MRK in out-of-sample covariance forecasts. We also note that the performances of QMLEs are not statistically distinguishable. In the second column, we report the results of the analysis from January 2009 to December 2011. The QMLE_{20} provides the lowest ex-post volatility, followed by $\text{QMLE}_{\hat{r}_{bic}}$, QMLE_{15}, QMLE_{10}, QMLE_5, and QMLE. The MRK and PCA provide higher ex-post volatility. As before, the p-value of QMLEs is significantly higher. Finally, in the last column, we report the results of the analysis from January 2012 to December 2014. The QMLE_{10} provides the lowest ex-post volatility, followed by QMLE_{20}, QMLE_{15}, $\text{QMLE}_{\hat{r}_{bic}}$, QMLE and QMLE_5. The relative performance of MRK and PCA is similar to the that in the second column. Furthermore, the MCS test again includes the QMLEs and excludes the PCA and MRK. Overall, we observe that, regardless of the time period, the QMLEs methods consistently outperform both PCA and MRK. We also observe that QMLE_i, $i = 5, 10, 15, 20$ are statistically indistinguishable from QMLE, so using QMLE or QMLE_i, $i = 5, 10, 15, 20$ is equivalent from a statistical viewpoint.

Table 3.3 The Out-of-sample portfolio construction.

Periods	$01/2006-12/2008$	$01/2009-12/2011$	$01/2012-12/2014$
Estimators	EV (p-value)	EV(p-value)	EV(p-value)
PCA	2.738(0.076)	1.863(0.000)	0.456(0.000)
MRK	2.446(0.076)	1.659(0.000)	0.664(0.000)
QMLE	**1.074***(1.000)	0.713*(0.102)	0.296*(0.403)
QMLE_{20}	1.081*(0.961)	**0.570***(1.000)	0.267*(0.978)
QMLE_{15}	1.202*(0.275)	0.608*(0.964)	0.274*(0.876)
QMLE_{10}	1.190*(0.275)	0.668*(0.495)	**0.266***(1.000)
QMLE_{5}	1.290*(0.117)	0.689*(0.331)	0.304*(0.119)
QMLE_{BIC}	1.226*(0.137)	0.606*(0.964)	0.278*(0.783)

Notes: Ex-post realized variances ($\times 10^4$) of GMV portfolios. The bold numbers highlighted the lowest loss function. The p-values of the MCS at 10% c.l. are indicated in parenthesis. The estimators included in the MCS are denoted with an asterisk

In Table 3.4 we present the results of the ex-post volatility and p-value of the MCS test in the entire sample from January 2006 to December 2014. Even in this case, all QMLEs outperform the PCA and MRK. In fact, the p-value of the MCS test is significantly larger, indicating that the QMLEs perform better with respect to out-of-sample covariance forecasts. The QMLE_{20} provides the lowest ex-post volatility, followed by QMLE, QMLE_{15}, $\text{QMLE}_{\hat{r}_{bic}}$ and QMLE_{10} , whereas the PCA provides the highest ex-post volatility, followed by MRK and QMLE_5. Overall, we conclude that the QMLEs provide the best out-of-sample covariance forecasts.

Table 3.4 The Out-of-sample portfolio construction.

	01/2006 - 12/2014
Estimator	EV(p-value)
PCA	5.061(0.005)
MRK	4.746(0.006)
QMLE	2.085*(0.293)
$QMLE_{20}$	**1.944***(1.000)
$QMLE_{15}$	2.085*(0.212)
$QMLE_{10}$	2.127*(0.212)
$QMLE_{5}$	2.283(0.031)
$QMLE_{BIC}$	2.106*(0.212)

Notes: The ex-post realized variances ($\times 10^4$) of GMV portfolios. The bold numbers highlighted the lowest loss function. The p-values of the MCS at 10% c.l. are indicated in parenthesis. The estimators included in the MCS are denoted with an asterisk

Computation gains

In Section (3.2), we highlighted the advantage of the realized factor QMLEs ($QMLE_{20}$, $QMLE_{15}$, $QMLE_{10}$, $QMLE_5$) compared to the QMLE. The main advantage is that the realized factor QMLE is computationally much simpler compared to the QMLE and is not affected by the curse of dimensionality. In this Section, we present the possible gain in computing time when evaluating the realized factor QMLEs. More precisely, we compare the average estimation time for the realized factor QMLEs and the average estimation time for the QMLE. The estimation times are average over 21 randomly selected trading days. Let $T_i, i = 5, 10, 15, 20, 100$ denote the average estimation time for the $QMLE_{20}$, $QMLE_{15}$, $QMLE_{10}$, $QMLE_5$ and QMLE, respectively.

In Table 3.5 we present the ratios between the average estimation time for the realized factor QMLEs and the average estimation time for the QMLE. The test is carried out on Asus VivoBook S (i7-7500U CPU at 2.90GHz). Results indicate that the $QMLE_5$ is the fastest method, followed by $QMLE_{10}$, $QMLE_{15}$ and $QMLE_{20}$, whereas the QMLE method is the slowest. Specifically, we find that the $QMLE_5$ method is 12 times faster than the QMLE method. The computational savings are substantial when the number of factor (r) is small, for example 5 or 10. If r is large, say, 20, the gains are less spectacular but nevertheless substantial.

Table 3.5 Computational gain.

i	$\frac{T_i}{T_{100}}$
5	**0.086**
10	0.126
15	0.192
20	0.313

Notes: The average estimation time for the realized factor QMLE and the average estimation time for the QMLE. The estimation times are average over 21 randomly trading days. The results are encouraging.

3.5.5 Value-at-Risk

Investor who invests in a portfolio at a particular point in time wants to determine the total loss of her portfolio at a given confidence level. To determine the total loss of the portfolio, most of the investors rely on the Value-at-Risk (VaR). The VaR is commonly used in the financial industry to manage the financial risk. Formally, *the VaR represents the maximum amount of money that may be lost on a portfolio over a given period of time, with given level of confidence* (Best 1998). The purpose of this section is to check whether the QMLE methods can correctly forecast the 1% and 5% VaR at 1-day horizon for the portfolios built in Section (3.5.4).

Let w_j be the portfolio weight at day j constructed in Section (3.5.4) from the 100 assets. The portfolio variance at day j, denote by $\hat{\sigma}_j^p$, is given by:

$$\hat{\sigma}_j^p = w_j^\top \hat{\Sigma}_j w_j \tag{3.46}$$

Let R_j be a portfolio return at day j, given by $R_j = w_j \Delta Y_j$ where ΔY is a vector of daily open-to-close returns. To compute the VaR, one can rely on three main approaches, namely the Variance-Covariance approach, Historical simulation and Monte Carlo simulation. In this study, we consider the Variance-Covariance method, which is widely used in financial applications. To use the variance-Covariance approach, we need to assume that the return of the assets follow the normal distribution with zero mean. Let ϕ_α^{-1} be the $\alpha-$quantile of the standard normal distribution and denote by α the coverage rate [6]. The VaR of the portfolio at day j with confidence level α, denoted by $\text{VaR}_j(\alpha)$, it is defined as:

$$\text{VaR}_j(\alpha) = \sqrt{\sigma_j^p}\,\phi_\alpha^{-1}. \tag{3.47}$$

[6]The coverage rate α is a probability that the lower tail VaR will be exceeded on a given day.

In practice, the coverage rate α is often set to be equal to 5%, 1%.

Backtesting Value-at-Risk

To assess how accurate the proposed model forecasts the VaR, we employ a standard backtesting procedure (see, Kupiec 1995 and Christoffersen 1998). Backtesting consists in comparing the portfolio losses to the estimated VaR. The most popular backtesting procedure is based on the *hit sequence* (or *violation process* as named by Christoffersen 1998). Let I_j be an indicator function, the hit sequence is defined as:

$$I_j(\alpha) = \begin{cases} 1 & \text{if } \mathrm{R}_j < \mathrm{VaR}_{j-1}(\alpha) \\ 0 & \text{if } \mathrm{R}_j \geq \mathrm{VaR}_{j-1}(\alpha) \end{cases}, j = 2, \ldots, T$$

where T is the total number of trading days. We can see that the hit sequence indicates whether the portfolio loss at day j is greater than VaR estimates. To backtest the proposed model, we adopt to the widely known test based of failure rates proposed by Kupiec (1995). In Kupiec (1995), the hit-sequence I_j is modeled as a sequence of i.i.d Bernoulli distributed variables:

$$I_j(\alpha) \sim Bernoulli(\bar{\alpha}), \quad j = 1, \ldots, T \tag{3.48}$$

where $\bar{\alpha} \in (0,1)$ is an unknown probability parameter. The likelihood of Bernoulli sequence (3.48) is given by $L(\bar{\alpha}) = \bar{\alpha}^{\tau^*}(1-\bar{\alpha})^{T-\tau^*}$, and the maximum likelihood of (3.48) is then given by

$$\bar{\alpha} = \frac{\tau^*}{T},$$

where $\tau^* = \sum_{j=1}^{T} \mathrm{I}_t$ is the total number of exception. One can see that the $\bar{\alpha}$ is the average number of exception or *failure rate*. The Kupiec's test (or proportion of failure test, POF-test) measures whether the amount of exceptions (or times the portfolio loss $\bar{\alpha}$ exceeds the estimated VaR) is consistent to the confidence level α. If the amount of exceptions is higher (lower) than the confidence level, this would indicate that the risk model underestimates (overestimates) the VaR. Therefore, under the null hypotheses, the total number of time the observed portfolio loss exceeds the estimated VaR will be equal to the confidence level.

The null hypothesis for the POF-test is as follows:

$$\mathrm{H}_0: \quad \bar{\alpha} = \alpha$$
$$\mathrm{H}_1: \quad \bar{\alpha} \neq \alpha.$$

To test the null hypothesis, we use the following likelihood ratio (LR) test proposed by Kupiec (1995):

$$\mathrm{LR}_K = 2\{-\log[\alpha^{\tau^*}(1-\alpha)^{T-\tau^*}] + \log[\bar{\alpha}^{\tau^*}(1-\bar{\alpha})^{T-\tau^*}]\}.$$

Under the null hypothesis, LR_K is asymptotically chi-squared distribution with one degree of freedom ($\chi^2(1)$). If the critical value of the LR_K exceeds the value of the $\chi^2(1)$ distribution, the null hypothesis will be rejected and therefore the model considered as incorrect. Bauwens and Laurent (2005) and Creal et al. (2011) used the similar statistic test to evaluate the performance of their models.

VaR analysis

Table 3.6 presents the results of the failure rate and the p-values of the Kupiec test for the periods of 2006-2008, 2009-20011 and 2012-2014 for the QMLEs, MRK and PCA. In the first column, we observe that the QMLE methods perform better overall. The observed failure rate ($\bar{\alpha}$) does not differ significantly from the expected failure rate (1%) for the VaR forecasts obtained with almost all QMLE methods. At the 5% level, the QMLE, QMLE_{20} and QMLE_{15} provide a better accuracy of VaR forecast than the QMLE_{10}, QMLE_5 and $\mathrm{QMLE}_{\mathrm{BIC}}$ according to the Kupiec test. In the second column, we observe that all QMLE methods perform better than PCA and MRK at the 1% confidence level. The p-values for rejecting the hypothesis of correct failure of 1% are significantly larger, implying that all QMLE methods generate valid VaR forecasts. At the 5% level, the MRK and all QMLE method provide a better accuracy of VaR forecast than the PCA according to the Kupiec test. The results in the last column indicate that all QMLE methods perform better than the MRK and PCA according to the Kupiec test.

Table 3.7 displays the results of the failure rate and the p-values for the Kupiec test for the full period from January 2006 to December 2014 for the QMLEs, MRK and PCA. Even when the sample size is large (9 years in this case), the QMLE methods still perform better at 1% level. In fact, we can see that the p-values for rejecting the hypothesis of correct failure are significant larger overall. At the 5% level, the QMLE_{20} and QMLE_{15} outperform the QMLE, QMLE_{10}, QMLE_5 and $\mathrm{QMLE}_{\mathrm{BIC}}$. Furthermore, we observe that the PCA and MRK perform worse in general. Overall, we conclude that

all QMLEs provide a better accuracy performance and in most of the cases the realized factor QMLE are statistically indistinguishable from the QMLE.

Table 3.6 One-step-ahead VaR backtesting at a 1% and 5% failure level for the 100 assets traded in NYSE within subsamples.

	$01/2006-12/2008$		$01/2009-12/2011$		$01/2012-12/2014$	
Estimators	1%	5%	1%	5%	1%	5%
PCA	5.31(0.000)	12.07(0.000)	2.25(0.003)	6.50(0.071)	3.59(0.000)	7.98(0.000)
MRK	3.18(0.000)	9.81(0.000)	1.98(0.016)	**5.42***(0.598)	3.99(0.000)	7.58(0.003)
QMLE	1.33*(0.391)	3.85*(0.130)	1.46*(0.239)	4.10*(0.242)	1.33*(0.387)	4.26*(0.337)
$QMLE_{20}$	1.33*(0.391)	**3.98***(0.183)	1.59*(0.135)	4.23*(0.321)	**1.06***(0.862)	4.65*(0.660)
$QMLE_{15}$	0.80*(0.559)	3.85*(0.130)	**1.19***(0.609)	4.50*(0.519)	1.99(0.0158)	**4.92***(0.920)
$QMLE_{10}$	**0.93***(0.841)	3.45(0.039)	0.66*(0.319)	4.23*(0.321)	1.20*(0.599)	4.65*(0.660)
$QMLE_5$	0.66*(0.322)	2.92(0.005)	1.46*(0.239)	5.82*(0.313)	1.20*(0.599)	3.86*(0.134)
$QMLE_{BIC}$	1.19*(0.604)	3.71(0.090)	**1.19***(0.609)	4.37*(0.413)	1.46*(0.233)	3.99*(0.188)

Notes: The average number of VaR exceedances α in percentage and p-values of the Kupiec test (in parentheses). Asterisks denote estimators for which the test is not rejected. The bold numbers highlight the model with the highest p-value.

Table 3.7 The results of VaR obtained in the whole sample from 2006 to 2014.

	$01/2006-12/2014$	
Estimator	1%	5%
PCA	3.76(0.000)	8.89(0.000)
MRK	3.01(0.000)	7.56(0.000)
QMLE	1.33*(0.138)	3.98(0.020)
$QMLE_{20}$	1.37(0.097)	4.33*(0.136)
$QMLE_{15}$	1.33*(0.137)	**4.42***(0.198)
$QMLE_{10}$	**0.93***(0.729)	4.11(0.046)
$QMLE_{5}$	1.11*(0.621)	4.20(0.073)
$QMLE_{BIC}$	1.28*(0.196)	4.02(0.027)

Notes: Frequencies of exceedances and *p*-values of the Kupiec test (in parentheses). Asterisks denote estimators for which the test is not rejected. The bold numbers highlight the model with the highest p-value.

3.6 Conclusion

In this work, we propose to model intraday prices through a dynamic factor model where idiosyncratic errors incorporate microstructure effects. Precisely, we propose a dynamic factor specification for log-prices rather than for log-returns. This modelling strategy allows as to easily handle the problem of data reduction since the common factors are filtered using all the available log-prices. The proposed model addresses the curse of dimensionality problem of the QMLE estimator of Corsi et al. (2015) and Shephard and Xiu (2017), and its estimation is performed by assuming a linear-Gaussian state-space representation and, employing an Expectation-Maximization algorithm. The model provides a fully parametric setup where one can treat asynchronicity as a missing data problem. The principal advantages are that covariances can be estimated using the entire informational content of the data and that the problem of model selection, namely how to select the correct number of factors, is easily addressed through a likelihood-based criterion. The advantage compared to QMLE is that the number of parameters scales well as the dimension increases.

The statistical properties of the realized factor QMLE are assessed through Monte Carlo simulations in which we showed that the proposed estimators remains unbiased under a misspecified DGP, as it correctly estimates the true covariance matrices in the presence of stochastic volatility in the factors and asynchronous trading. Finally, in the empirical application, we study the factor structure of a cross-sectional of 100 NYSE assets and perform an horse race based on out-of-sample minimum variance portfolios to compare our estimator to PCA, MRK and QMLEs. We find that the factor

structure emerges at the sampling frequency of one minute. This is likely to be due the existence lead-lag dependencies at higher frequencies that cannot be captured by the model. Our empirical results show that the proposed estimators outperform the two benchmark methods (PCA and MRK) regarding out-of-sample ex-post forecasts, and at same time perform as good as the QMLE. However, they are computational much simpler. For instance, we find that QMLE_5 is 12 times faster than QMLE on average. Lastly, the results show that the realized factor QMLE and QMLE provide a better forecast of 1% and 5% VaR compared to PCA and MRK. The performance of the realized factor QMLE are not statistically distinguishable from the QMLE, so using QMLE or QMLE_i, $i = 5, 10, 15, 20$ and BIC is equivalent.

B Appendix

B.1 Kalman smoother, Kalman filter and lag-1 covariance smoother algorithms

The Kalman filter presented below is a same as in (Shumway and Stoffer (2010), pg $296 - 304$) but with the little modification due to the model.

$$G_t^{t-1} = G_{t-1}^{t-1} \tag{B.1}$$

$$V_t^{t-1} = V_{t-1}^{t-1} + \Lambda_{tr} \tag{B.2}$$

$$G_t^t = G_t^{t-1} + K_t(Y_t - H_{rt}G_t^{t-1}) \tag{B.3}$$

$$V_t^t = V_t^{t-1} - K_t H_{rt} V_t^{t-1} \tag{B.4}$$

$$K_t = V_t^{t-1}(H_{rt})^\top (H_{rt}V_t^{t-1}H^\top + R_t)^{-1}. \tag{B.5}$$

Note that the Kalman filter differs to the kalman smoother since, the filter one also provides the expectation of G_t conditioned on the data up to time t but not on the entire data (i.e. from time 1 to d). Then, the Kalman filter is only the first part of the Kalman smooth algorithm since the filter provides the expectation of G_t at time t conditioned on the data up to time t. Then the Kalman smoother in which the expectation is conditioned on the entire data (i.e. from the time 1 to d) is given:

$$G_{t-1}^n = G_{t-1}^{t-1} + J_{t-1}(\tilde{G}_t - G_t^{t-1}) \tag{B.6}$$

$$V_{t-1}^n = V_{t-1}^{t-1} + J_{t-1}(\tilde{V}_t - V_t^{t-1})J_{t-1}^\top \tag{B.7}$$

$$J_{t-1} = V_{t-1}^{t-1}(V_t^{t-1})^{-1}, \tag{B.8}$$

and the $lag-1$ covariance smoother in which also the expectation is conditioned on the entire data (i.e. from the time 1 to d) is given:

$$V^n_{T,T-1} = (I - K_T H_T) V^n_{T-1} \tag{B.9}$$

$$V^T_{t-1,t-2} = V^{t-1}_{t-1} J^\top_{t-2} + J_{t-1}(V^T_{t,t-1} - V^{t-1}_{t-1}) J^\top_{t-2}. \tag{B.10}$$

Note that in the above algorithms when we are estimated the expectations conditioned in the observed data Y_t, ones are not taking into the consideration the presence of the missing components in the observed processes.

B.2 Proof of the propositions 6 and 7

Let us stech the proof of the propositions 6 and 7.

$$\frac{\partial \mathrm{Tr}\{\sum_{t=1}^n \left((Y_t - HG^n_t)(Y_t - HG^n_t)^\top + HV^n_t H^\top\right)\}}{R\partial H} = -R^{-1}\sum_{t=1}^n (G^n_t (Y_t - H^{j+1} G^n_t)^\top + V^n_t (H^{(j+1)})^\top).$$

Taking

$$\left.\frac{\partial F(\Omega|\mathscr{Y}_n;\Omega^j)}{\partial H}\right|_{H=H^{j+1}} = 0,$$

and yields the following:

$$\begin{aligned}
\sum_{t=1}^n G^n_t (Y_t - H^{j+1} G^n_t)^\top &= \sum_{t=1}^n V^n_t (H^{(j+1)})^\top \\
\sum_{t=1}^n G^n_t Y_t^\top - \sum_{t=1}^n G^n_t (G^n_t)^\top (H^{j+1})^\top &= \sum_{t=1}^n V^n_t (H^{(j+1)})^\top \\
\sum_{t=1}^n G^n_t Y_t^\top &= \sum_{t=1}^n (V^n_t + G^n_t (G^n_t)^\top)(H^{(j+1)})^\top.
\end{aligned} \tag{B.11}$$

B Appendix

Taking the transpose in both side of (B.11) and also using the fact that $(V_t^T)^\top = V_t^T$, thus the update parameter is given by

$$H^{j+1} = \sum_{t=1}^{n}(Y_t(G_t^T)^\top)\Big(\sum_{t=1}^{n}(V_t^T + G_t^n(G_t^T)^\top)\Big)^{-1}. \tag{B.12}$$

In the similar way as in the case of the update parameter of H, we get

$$\frac{\partial F(\Omega|\mathscr{Y}_n;\Omega^j)}{\partial R} = \frac{T}{R} - \sum_{t=1}^{n}\Big(\frac{(Y_t - HG_t^n)(Y_t - HG_t^n)^\top + HV_t^nH^\top}{R^2}\Big).$$

Taking

$$\frac{\partial F(\Omega|\mathscr{Y}_n;\Omega^j)}{\partial R}\bigg|_{R=R^{j+1}} = 0,$$

and we obtain

$$R^{j+1} = \sum_{t=1}^{n}\Big(\frac{(Y_t - HG_t^n)(Y_t - HG_t^n)^\top + HV_t^nH^\top}{T}\Big).$$

Then, the update parameter is given by

$$R^{j+1} = \text{diag}\left(\frac{1}{T}\sum_{t=1}^{n}(Y_t - H^{j+1}G_t^n)(Y_t - H^{j+1}G_t^n)^\top + H^{j+1}V_t^nH^{(j+1)T}\right). \tag{B.13}$$

Similarly as in the previous case and we also get

$$\frac{\partial F(\Omega|\mathscr{Y}_n;\Omega^j)}{\partial \Lambda} = \frac{T}{\Lambda} - \Big(\frac{S_{11} - S_{10} - S_{10}^\top + S_{00}}{\Lambda^2}\Big).$$

Taking

$$\frac{\partial F(\Omega|\mathscr{Y}_n;\Omega^j)}{\partial \Lambda}\bigg|_{\Lambda=\Lambda^{j+1}} = 0,$$

and the update parameter is given by

$$\Lambda_r^{j+1} = \frac{1}{T}\Big(S_{11} - S_{10} - (S_{10})^\top + S_{00}\Big). \tag{B.14}$$

B.3 Proof of the Proposition 8

Let us sketch the proof of the Proposition (8). The similar proof can be found in Bańbura and Modugno (2014). Suppose that at time t some of the observations in Y_t are missing. In this case, one needs to modify the conditional expectation of (3.4) in order to keep into account the missing observations. Shumway and Stoffer (1982), Harvey (1990) and Durbin and Koopman (2012) show how to modify the for Kalman filtering and smoothing recursions in the case of missing values.

Assume that A_t is a $d_t \times d$ matrix where d_t represents the number of observations available at time t. When we take the conditional expectation in Eq. (3.4), we only interest on the third term of (3.4) since the expectation of the first two terms in Eq. (3.4) depend only on G_t^n, V_t^n and $V_{t,t-1}^T$. Employ the E-step [7] and M-step as in the preview subsection, the update factor loading formula in the case of the missing values is given as follows: suppose that $\Phi_n \subseteq Y_n$ is a set of the observe prices, and $\Phi_n = Y_n$ if there is no missing values in Y_t. More precisely, Φ_n contains the observe component of Y_t. Because of the law of iterative expectation, the conditional expectation of the second term in Eq. (3.4) can be written as:

$$E\big((Y_t - HG_t)(Y_t - HG_t)^\top|\Phi_n\big) = E\big[\{E\big((Y_t - HG_t)(Y_t - HG_t)^\top|\Phi_n, G_t\big)\}|\Phi_n\big]. \tag{B.15}$$

We need to compute $E[(Y_t - HG_t)(Y_t - HG_t)^\top|\Phi_n, G_t]$. Let us first expand the term $(Y_t - HG_t)(Y_t - HG_t)^\top$ as:

[7] In the E-step we take expectation of (3.4) condition on the set of observe data Φ_n instead on Y_n since the vector Y_n contains the missing observations (see Shumway and Stoffer 1982)

B Appendix

$$
\begin{aligned}
(Y_t - HG_t)(Y_t - HG_t)^\top &= (A_t(Y_t - HG_t) + (I_d - A_t)(Y_t - HG_t))(A_t(Y_t - HG_t) + (I_d - A_t)(Y_t - HG_t))^\top \\
&= A_t(Y_t - HG_t)(Y_t - HG_t)^\top A_t + A_t(Y_t - HG_t)(Y_t - HG_t)^\top (I_d - A_t) \\
&+ (I_d - A_t)(Y_t - HG_t)(Y_t - HG_t)^\top A_t \\
&+ (I_d - A_t)(Y_t - HG_t)(Y_t - HG_t)^\top (I_d - A_t), \qquad \text{(B.16)}
\end{aligned}
$$

where $(I_d - A_t)Y_t$ contains no-missing observations at time t and 0 in the place of the missing values. The matrix I_d is $d \times d$ identity matrix. We have

$$
\begin{aligned}
E[(Y_t - HG_t)(Y_t - HG_t)^\top | \Phi_n, G_t] &= E[A_t(Y_t - HG_t)(Y_t - HG_t)^\top A_t | \Phi_n, G_t] \\
&+ E[A_t(Y_t - HG_t)(Y_t - E[HG_t)^\top (I_d - A_t) | \Phi_n, G_t] \\
&+ E[(I_d - A_t)(Y_t - HG_t)(Y_t - HG_t)^\top A_t | \Phi_n, G_t] \\
&+ E[(I_d - A_t)(Y_t - HG_t)(Y_t - HG_t)^\top (I_d - A_t) | \Phi_n, G_t]. \qquad \text{(B.17)}
\end{aligned}
$$

Let us now evaluate each of the term at the right-side of Eq. (B.17). Due to the fact that Y_t follow the multivariate normal distribution, and by applying the properties of conditional probability of the multivariate normal distribution [8], we get

$$
\begin{aligned}
E[A_t(Y_t - HG_t)(Y_t - HG_t)^\top (I_d - A_t) | \Phi_n, G_t] &= cov_{\Phi|G}[A_t\eta_t, (I_d - A_t)\eta_t | \Phi_n, G_t] \\
&= 0 \qquad \text{(B.18)} \\
E[(I_d - A_t)(Y_t - HG_t)(Y_t - HG_t)^\top (I_d - A_t) | \Phi_n, G_t] &= var_{\Phi|G}[(I_d - A_t)\eta_t | \Phi_n, G_t] \\
&+ E[(I_d - A_t)\eta_t | Y_n, G_t] \\
&\times (E[(I_d - A_t)\eta_t | \Phi_n, G_t])^\top \\
&= (I_d - A_t)R^j(I_d - A_t), \qquad \text{(B.19)}
\end{aligned}
$$

where $cov_{\Phi|G}$ and $var_{\Phi|G}$ mean conditional covariance and variance. We then get

[8] see for example Lipsitz et al. 2000

$$\begin{aligned}
E[E[A_t(Y_t - HG_t)(Y_t - HG_t)^\top A_t|\Phi_n, G_t]|Y_n] =& E[A_t(Y_t - HG_t)(Y_t - HG_t)^\top A_t|\Phi_n] \\
=& E[\{A_tY_tY_t^\top A_t - A_tY_t(HG_t)^\top A_t - A_tHG_tY_t^\top A_t \\
&+ A_tHG_tG_t^\top H^\top A_t\}|\Phi_n] \\
=& A_tY_tY_t^\top A_t - A_tY_t(G_t^n)^\top H^\top A_t - A_tHG_t^nY_t^\top A_t \\
&+ A_tH(V_t^n + G_t^n(G_t^n)^\top)H^\top A_t \\
=& (A_tY_t - A_tHG_t^n)(A_tY_t - A_tHG_t^n)^\top + A_tHV_t^nH^\top A_t \qquad \text{(B.20)}
\end{aligned}$$

Combining the results (B.18), (B.19) and (B.20) to obtain

$$\begin{aligned}
E[(Y_t - HG_t)(Y_t - HG_t)^\top|\Phi_n] =& A_tY_t(A_tY_t)^\top - A_tY_t(G_t^n)^\top H^\top A_t - A_tHG_t^n(A_tY_t)^\top \\
&+ A_tH(V_t^n + G_t^n(G_t^n)^\top H^\top A_t + (I_d - A_t)R^j(I_d - A_t) \\
=& (A_tY_t - A_tHG_t^n)(A_tY_t - A_tHG_t^n)^\top + A_tHV_t^nH^\top A_t \\
&+ (I_d - A_t)R^j(I_d - A_t). \qquad \text{(B.21)}
\end{aligned}$$

Substituting (B.21) in (3.4) and employ the E-step and M-step as in the preview subsection to obtain

$$\frac{\partial F(\Omega|\mathscr{Y}_n;\Omega^j)}{\partial H} = \sum_{t=1}^{n} R^{-1}(-A_tG_t^n(A_tY_t - A_tHG_t^n)^\top + A_tV_t^n(A_tH)^\top). \qquad \text{(B.22)}$$

Taking

$$\left.\frac{\partial F(\Omega|\mathscr{Y}_n;\Omega^j)}{\partial H}\right|_{H=H^{j+1}} = 0,$$

we get

$$\sum_{t=1}^{n} G_t^n (A_t Y_t - A_t H^{j+1} G_t^n)^\top = \sum_{t=1}^{n} V_t^n (A_t H^{(j+1)})^\top$$
$$\sum_{t=1}^{n} G_t^n (A_t Y_t)^\top - \sum_{t=1}^{n} G_t^n (G_t^n)^\top (A_t H^{j+1})^\top = \sum_{t=1}^{n} V_t^n (A_t H^{(j+1)})^\top$$
$$\sum_{t=1}^{n} G_t^n (A_t Y_t)^\top = \sum_{t=1}^{n} (V_t^n + G_t^n (G_t^n)^\top)(A_t H^{(j+1)})^\top. \tag{B.23}$$

Taking the transpose in both side in Eq. (B.23) and also using the fact that $(V_t^n)^\top = V_t^n$, we get

$$\sum_{t=1}^{n} A_t Y_t (G_t^n)^\top = \sum_{t=1}^{n} A_t H^{(j+1)} (V_t^n + G_t^n (G_t^n)^\top).$$

It well-known that $\text{vec}(EDC) = (C' \otimes E)\text{vec}(D)$ and thus by vectorizing the preview equation we get

$$\text{vec}\Big(\sum_{t=1}^{n} (I - A_t) Y_t (G_t^n)^\top\Big) = \Big(\sum_{t=1}^{n} (V_t^n + G_t^n (G_t^n)^\top) \otimes A_t\Big)\text{vec}(H^{(j+1)}). \tag{B.24}$$

Dividing (B.24) by $\Big(\sum_{t=1}^{n} (V_t^n + G_t^n (G_t^n)^\top) \otimes A_t\Big)$ to obtain

$$\text{vec}(H^{(j+1)}) = \Big(\sum_{t=1}^{n} (V_t^n + G_t^n (G_t^n)^\top) \otimes A_t\Big))^{-1} \text{vec}\Big(\sum_{t=1}^{n} (I - A_t) Y_t (G_t^n)^\top\Big). \tag{B.25}$$

Similarly as in the previous case, we get

$$\frac{\partial F(\Omega|\Phi_n;\Omega^j)}{\partial R} = R^{-1} n - \sum_{t=1}^{n} \Big((A_t Y_t - A_t H G_t^n)(A_t Y_t - A_t H G_t^n)^\top + A_t H V_t^n H^\top A_t + (I_d - A_t) R^j (I_d - A_t)\Big) R^{-2}.$$

Taking

$$\frac{\partial F(\Omega|\mathscr{Y}_n;\Omega^j)}{\partial R}\bigg|_{R=R^{j+1}} = 0,$$

we get

$$(R^{j+1})^{-1}n = R^{-2(j+1)}\sum_{t=1}^{n}\Big((A_tY_t - A_tH^{j+1}G_t^n)(A_tY_t - A_tH^{j+1}G_t^n)^\top + A_tH^{j+1}V_t^n(H^{j+1})^\top A_t + (I_d - A_t)R^j(I_d - A_t)\Big).$$

Thus, the update R is given by:

$$R^{j+1} = \text{diag}\left(\frac{1}{n}\sum_{t=1}^{n}\left\{(A_tY_t - A_tH^{j+1}G_t^n)(A_tY_t - A_tH^{j+1}G_t^n)^\top + A_tH^{j+1}V_t^n(H^{j+1})^\top A_t + (I_d - A_t)R^j(I_d - A_t)\right\}\right). \tag{B.26}$$

B.4 Summary on EM algorithm

Here we summarize how to implement the EM algorithm for dynamic factor model with missing value. As suggested by Shumway and Stoffer (1982) that both G_0 and V_0 cannot be estimated simultaneously with the EM algorithm, therefore one needs to keep fix V_0 and estimate G_0 within the EM algorithm. We follow their suggestion by keeping the diagonal matrix V_0 fixes with all diagonal elements are set to be 10^5. The implementation of EM algorithm is described as follows:

1. Start by choosing the initial values of R, Λ, H.
2. For k=0,1,2, $\cdots$
   - E-step: estimate $G_t^{T(k)}$ and $V_t^{T(k)}$ by Kalman smoothing (B.1).
   - M-Step: update the parameters Λ, R and H as given in Eq. (3.12), (3.13) and (3.11)
   - Stop on convergence when change in $\log L < 10^{-6}$. The $\log L$ is given by (3.14)

Bibliography

M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1965.

Y. Ait-Sahalia. Nonparametric pricing of interest rate derivative securities. Technical report, National Bureau of Economic Research, 1995.

Y. Ait-Sahalia. Testing continuous-time models of the spot interest rate. *The review of financial studies*, 9(2):385–426, 1996.

Y. Ait-Sahalia and D. Xiu. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399, 2017.

Y. Ait-Sahalia and D. Xiu. Principal component analysis of high-frequency data. *Journal of the American Statistical Association*, 114(525):287–303, 2019.

H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974a.

H. Akaike. A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pages 215–222. Springer, 1974b.

T. G. Andersen, O. Bondarenko, A. S. Kyle, and A. A. Obizhaeva. Intraday trading invariance in the e-mini s&p 500 futures market. 2016.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

J. Bai and P. Wang. Identification and bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics*, 33(2):221–240, 2015.

V. Bally and D. Talay. The law of the euler scheme for stochastic differential equations. *Probability theory and related fields*, 104(1):43–60, 1996.

M. Bańbura and M. Modugno. Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160, 2014.

F. Bandi and P. C. Phillips. Nonstationary continuous-time processes. *Handbook of financial econometrics*, 1:139–202, 2009.

F. M. Bandi. Short-term interest rate dynamics: a spatial approach. *Journal of Financial Economics*, 65(1):73–110, 2002.

F. M. Bandi and P. C. Phillips. Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71(1):241–283, 2003.

G. Barlevy and P. Veronesi. Rational panics and stock market crashes. *Journal of Economic Theory*, 110(2):234–263, 2003.

O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925, 2004.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169, 2011.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Realized kernels in practice: trades and quotes. *The Econometrics Journal*, 12(3):C1–C32, 11 2009. ISSN 1368-4221. doi: 10.1111/j.1368-423X.2008.00275.x. URL https://doi.org/10.1111/j.1368-423X.2008.00275.x.

L. Bauwens and S. Laurent. A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics*, 23(3):346–354, 2005.

M. Baxter and A. Ronnie. *Financial calculus: an introduction to derivative rpicing*. Cambridge University Press, 1996.

M. Bayram, T. Partal, and G. O. Buyukoz. Numerical methods for simulation of stochastic differential equations. *Advances in Difference Equations*, 2018(1):17, 2018.

P. Best. *Implementing value at risk*. Wiley Online Library, 1998.

B. M. Bibby and M. Sørensen. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, pages 17–39, 1995.

B. Billah, R. J. Hyndman, and A. B. Koehler. Empirical information criteria for time series forecasting model selection. *Journal of Statistical Computation and Simulation*, 75(10):831–840, 2005.

F. Black. Capital market equilibrium with restricted borrowing. *The journal of business*, 45(3):444–455, 1972.

T. Bollerslev and B. Y. Zhang. Measuring and modeling systematic risk in factor pricing models using high-frequency data. *Journal of Empirical Finance*, 10(5):533–558, 2003.

L. Bork. Estimating us monetary policy shocks using a factor-augmented vector autoregression: An em algorithm approach. *Available at SSRN 1348552*, 2009.

L. Bork, H. Dewachter, and R. Houssa. Identification of macroeconomic factors in large panels. *CREATES Research Paper*, (2009-43), 2009.

D. R. Brillinger. Modeling spatial trajectories. *Handbook of spatial statistics*, pages 463–475, 2010.

M. K. Brunnermeier and L. H. Pedersen. Predatory trading. *The Journal of Finance*, 60(4):1825–1863, 2005.

G. Buccheri, G. Livieri, D. Pirino, and A. Pollastri. A closed-formula characterization of the epps effect. *Available at SSRN 3255070*, 2018.

G. Buccheri, G. Bormetti, F. Corsi, and F. Lillo. A score-driven conditional correlation model for noisy and asynchronous data: An application to high-frequency covariance dynamics. *Available at SSRN 2912438*, 2019.

K. Bulteel, T. F. Wilderjans, F. Tuerlinckx, and E. Ceulemans. Chull as an alternative to aic and bic in the context of mixtures of factor analyzers. *Behavior Research Methods*, 45(3):782–791, 2013.

T. Cai and W. Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011.

M. Carletti, K. Burrage, and P. M. Burrage. Numerical simulation of stochastic ordinary differential equations in biomathematical modelling. *Mathematics and Computers in Simulation*, 64(2):271–277, 2004.

J. E. Cavanaugh and R. H. Shumway. A bootstrap variant of aic for state-space model selection. *Statistica Sinica*, pages 473–496, 1997.

E. Ceulemans and I. Van Mechelen. Hierarchical classes models for three-way three-mode binary data: Interrelations and model selection. *Psychometrika*, 70(3):461–480, 2005.

S. CFTC and U. SEC. Findings regarding the market events of may 6, 2010. *Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*, 2010.

K. C. Chan, G. A. Karolyi, F. A. Longstaff, and A. B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *The journal of finance*, 47(3):1209–1227, 1992.

K. Christensen, R. C. Oomen, and R. Renò. The drift burst hypothesis. 2016.

P. F. Christoffersen. Evaluating interval forecasts. *International economic review*, pages 841–862, 1998.

R. Cont and L. Wagalath. Running for the exit: distressed selling and endogenous correlation in financial markets. *Mathematical Finance*, 23(4):718–741, 2013.

R. Cont and L. Wagalath. Fire sales forensics: measuring endogenous risk. *Mathematical Finance*, 26(4):835–866, 2016.

F. Corsi, S. Peluso, and F. Audrino. Missing in asynchronicity: A kalman-em approach for multivariate realized covariance estimation. *Journal of Applied Econometrics*, 30(3):377–397, 2015.

J. C. Cox, J. E. Ingersoll Jr, and S. A. Ross. A theory of the term structure of interest rates. In *Theory of Valuation*, pages 129–164. World Scientific, 2005.

D. Creal, S. J. Koopman, and A. Lucas. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics*, 29(4):552–563, 2011.

D. Creal, B. Schwaab, S. J. Koopman, and A. Lucas. Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *The Review of Economics and Statistics*, 96(5):898–915, 2014.

J. Danielsson, H. S. Shin, and J.-P. Zigrand. Endogenous extreme events and the dual role of prices. *Annu. Rev. Econ.*, 4(1): 111–129, 2012.

J. Davidson. *Stochastic limit theory: An introduction for econometricians*. OUP Oxford, 1994.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

C. Doz, D. Giannone, and L. Reichlin. A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164(1):188 – 205, 2011. ISSN 0304-4076. Annals Issue on Forecasting.

C. Doz, D. Giannone, and L. Reichlin. A quasi–maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics*, 94(4):1014–1024, 2012.

D. Duffie and K. J. Singleton. Simulated moments estimation of markov models of asset prices, 1990.

J. Durbin and S. J. Koopman. *Time series analysis by state space methods*, volume 38. OUP Oxford, 2012.

G. B. Durham and A. R. Gallant. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20(3):297–338, 2002.

H. Engelbert and W. Schmidt. Strong markov continuous local martingales and solutions of one-dimensional stochastic differential equations (part iii). *Mathematische Nachrichten*, 151(1):149–197, 1991.

H. J. Engelbert and W. Schmidt. On one-dimensional stochastic differential equations with generalized drift. In *Stochastic Differential Systems Filtering and Control*, pages 143–155. Springer, 1985.

R. Engle and R. Colacito. Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, 24(2):238–253, 2006.

T. W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366a): 291–298, 1979.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33 (1):3–56, 1993.

J. Fan, Y. Liao, and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680, 2013.

J. Fan, Y. Liao, and H. Liu. An overview of the estimation of large covariance and precision matrices, 2016.

D. Florens-Zmirou. Approximate discrete-time schemes for statistics of diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 20(4):547–557, 1989.

D. Florens-Zmirou. On estimating the diffusion coefficient from discrete observations. *Journal of applied probability*, 30 (4):790–804, 1993.

A. R. Gallant. Testing a nonlinear regression specification: A nonregular case. *Journal of the American Statistical Association*, 72(359):523–530, 1977.

G. Gennotte and H. Leland. Market liquidity, hedging, and crashes. *The American Economic Review*, pages 999–1021, 1990.

J. Geweke. The dynamic factor analysis of economic time series. *Latent Variables in Socio-Economic Models*, 1977.

J. F. Geweke and K. J. Singleton. Maximum likelihood" confirmatory" factor analysis of economic time series. *International Economic Review*, pages 37–54, 1981.

D. Giannone, L. Reichlin, and L. Sala. Monetary policy in real time. Working Papers 284, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University, 0 2005.

D. Giannone, L. Reichlin, and D. Small. Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676, May 2008.

C. Gourieroux, A. Monfort, and E. Renault. Indirect inference. *Journal of applied econometrics*, 8(S1):S85–S118, 1993.

J. E. Griffin and R. C. Oomen. Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *Journal of Econometrics*, 160(1):58–68, 2011.

S. J. Grossman and M. H. Miller. Liquidity and market structure. *the Journal of Finance*, 43(3):617–633, 1988.

J. Hale. *Ordinary differential equations*. Pure and applied mathematics. Wiley-Interscience, 1969. URL https://books.google.it/books?id=2jOxAAAAIAAJ.

P. R. Hansen, A. Lunde, and J. M. Nason. The model confidence set. *Econometrica*, 79(2):453–497, 2011.

W. Härdle. *Applied nonparametric regression*. Number 19. Cambridge University Press, 1990.

A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.

N. Hautsch, L. M. Kyj, and R. C. Oomen. A blocking and regularization approach to high-dimensional realized covariance estimation. *Journal of Applied Econometrics*, 27(4):625–645, 2012.

T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11 (2):359–379, 2005.

D. J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review*, 43 (3):525–546, 2001.

N. Ikeda and S. Watanabe. Stochastic differential equations and diffusion processes. 1989.

J. I. Jeisman. *Estimation of the parameters of stochastic differential equations*. PhD thesis, Queensland University of Technology, 2006.

Z. Jian, P. Deng, and Z. Zhu. High-dimensional covariance forecasting based on principal component analysis of high-frequency data. *Economic Modelling*, 75:422–431, 2018.

G. J. Jiang. Nonparametric modeling of us interest rate term structure dynamics and implications on the prices of derivative securities. *Journal of financial and quantitative analysis*, 33(4):465–497, 1998.

G. J. Jiang and J. L. Knight. A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econometric Theory*, 13(5):615–645, 1997.

I. Karatzas and J. Ruf. Distribution of the time to explosion for one-dimensional diffusions. *Probability Theory and Related Fields*, 164(3-4):1027–1069, 2016.

I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24(2): 211–229, 1997.

A. Kirilenko, A. S. Kyle, M. Samadi, and T. Tuzun. The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3):967–998, 2017.

F. B. Knight. A reduction of continuous square-integrable martingales to brownian motion. In *Martingales*, pages 19–31. Springer, 1971.

S. J. Koopman et al. Likelihood-based analysis for dynamic factor models. 2008.

P. Kupiec. Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives*, 3(2), 1995.

G. Leobacher, M. Szölgyenyi, S. Thonhauser, et al. On the existence of solutions of a class of sdes with discontinuous drift and singular diffusion. *Electronic Communications in Probability*, 20, 2015.

J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In *Stochastic optimization models in finance*, pages 131–155. Elsevier, 1975.

S. R. Lipsitz, G. Molenberghs, G. M. Fitzmaurice, and J. Ibrahim. Gee with gaussian estimation of the correlations when data are incomplete. *Biometrics*, 56(2):528–536, 2000.

R. S. Liptser and A. N. Shiryayev. Theory of martingales, vol. 49. *Mathematics and its Applications (Soviet Series)*, 1989.

M. E. Mancino and S. Sanfelici. Estimating covariance via fourier method in the presence of asynchronous trading and microstructure noise. *Journal of financial econometrics*, 9(2):367–408, 2011.

X. Mao. *Stochastic differential equations and applications*. Elsevier, 2007.

X. Mao and C. Yuan. *Stochastic differential equations with Markovian switching*. Imperial College Press, 2006.

H. Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

G. Maruyama. Continuous markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4 (1):48–90, 1955.

S. Morris and H. S. Shin. Liquidity black holes. *Review of Finance*, 8(1):1–18, 2004.

J. Nicolau. Bias reduction in nonparametric diffusion coefficient estimation. *Econometric Theory*, 19(5):754–777, 2003a.

J. Nicolau. Bias reduction in nonparametric diffusion coefficient estimation. *Econometric Theory*, 19(5):754–777, 2003b.

K. B. Nowman. Gaussian estimation of single-factor continuous time models of the term structure of interest rates. *The journal of Finance*, 52(4):1695–1706, 1997.

B. Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.

T. Ozaki. A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. *Statistica Sinica*, pages 113–135, 1992.

E. Pardoux and D. Talay. Discretization and simulation of stochastic differential equations. *Acta Applicandae Mathematica*, 3(1):23–47, 1985.

A. J. Patton and K. Sheppard. Evaluating volatility and correlation forecasts. In *Handbook of financial time series*, pages 801–838. Springer, 2009.

M. Pelger. Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics*, 208(1): 23–42, 2019a.

M. Pelger. Understanding systematic risk: A high-frequency approach. *Available at SSRN 2647040*, 2019b.

E. Platen. An introduction to numerical methods for stochastic differential equations. *Acta numerica*, 8:197–246, 1999.

B. Prakasa Rao. Statistical inference for diffusion type processes. *Kendall's Lib. Statist.*, 8, 1999.

H. K. Preisler, A. A. Ager, and M. J. Wisdom. Analyzing animal movement patterns using potential functions. *Ecosphere*, 4(3):1–13, 2013.

T. Proietti. Estimation of common factors under cross-sectional and temporal aggregation constraints: nowcasting monthly gdp and its main components. In *COMPSTAT 2008*, pages 547–558. Springer, 2008.

R. Renò. A closer look at the epps effect. *International Journal of Theoretical and Applied Finance*, 6(01):87–102, 2003.

R. Renò. Nonparametric estimation of the diffusion coefficient of stochastic volatility models. *Econometric Theory*, 24(5): 1174–1206, 2008.

D. Revuz and M. Yor. Continuous martingales and brownian motion, vol. 293 of grundlehren der mathematischen wissenschaften [fundamental principles of mathematical sciences], 1999.

A. J. Rothman, P. J. Bickel, E. Levina, J. Zhu, et al. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515, 2008.

S. Rubenthaler. Numerical simulation of the solution of a stochastic differential equation driven by a lévy process. *Stochastic processes and their applications*, 103(2):311–349, 2003.

W. Rudin. Principles of mathematical analysis (international series in pure & applied mathematics). 1976.

Y. Saito and T. Mitsui. Simulation of stochastic differential equations. *Annals of the Institute of Statistical Mathematics*, 45 (3):419–432, 1993.

T. J. Sargent and C. A. Sims. Business cycle modeling without pretending to have too much a priori economic theory. Working Papers 55, Federal Reserve Bank of Minneapolis, 1977.

T. Sauer. Numerical solution of stochastic differential equations in finance. In *Handbook of computational finance*, pages 529–550. Springer, 2012.

G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3): 425–442, 1964.

N. Shephard and D. Xiu. Econometric analysis of multivariate realised qml: Estimation of the covariation of equity prices under asynchronous trading. *Journal of Econometrics*, 201(1):19–42, 2017.

I. Shoji and T. Ozaki. Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications*, 16(4):733–752, 1998.

R. Shumway. *Applied time series abalysis*. na, 1988.

R. H. Shumway and D. S. Stoffer. An approach to time series smoothing and forecasting using the em algorithm. *Journal of time series analysis*, 3(4):253–264, 1982.

R. H. Shumway and D. S. Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2010.

R. Stanton. A nonparametric model of term structure dynamics and the market price of interest rate risk. *The journal of Finance*, 52(5), 1997.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179, 2002.

D. Talay. Simulation and numerical analysis of stochastic differential systems: a review. 1990.

N. M. Temme. *Special functions: An introduction to the classical functions of mathematical physics*. John Wiley & Sons, 2011.

A. Tocino and R. Ardanuy. Runge–kutta methods for numerical solution of stochastic differential equations. *Journal of Computational and Applied Mathematics*, 138(2):219–241, 2002.

A. J. Veretennikov. On strong solutions and explicit formulas for solutions of stochastic integral equations. *Sbornik: Mathematics*, 39(3):387–403, 1981.

M. W. Watson and R. F. Engle. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400, 1983.

C. J. Wu et al. On the convergence properties of the em algorithm. *The Annals of statistics*, 11(1):95–103, 1983.

T. Yamada, S. Watanabe, et al. On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, 11(1):155–167, 1971.

N. Yoshida. Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41(2):220–242, 1992.

L. Zhang. Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47, 2011.

A. K. Zvonkin. A transformation of the phase space of a diffusion process that removes the drift. *Mathematics of the USSR-Sbornik*, 22(1):129, 1974.